DOCUMENT RESUME

ED 277 399	IR 051 742

AUTHOR	Katzer, Jeffrey; And Others
TITLE	Impact of Anaphoric Resolution in Information Retrieval. Final Report.
INSTITUTION	Syracuse Univ., N.Y. School of Information Studies.
SPONS AGENCY	National Science Foundation. Washington, D.C. Div. of Information Science and Technology.
PUB DATE	Oct 86
GRANT	NSF-IST-8313716
NOTE	369p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC15 Plus Postage.
DESCRIPTORS	*Abstracts; Algorithms; Classification; Comparative Analysis; Discourse Analysis; *Information Retrieval; *Relevance (Information Retrieval); Statistical Analysis; Tables (Data); *User Satisfaction (Information)
IDENTIFIERS	*Anaphora; *Automatic Content Analysis; Bibliographic Data Bases; Frequency Analysis; INSPEC; Linguistic Analysis; Psychological Abstracts; Referents (Linguistics); Weighted Term Searching

ABSTRACT

This project examines anaphora (the linguistic device of abbreviated subsequent reference to a concept) in information retrieval (IR) systems in order to develop procedures to recognize anaphors in text and distinguish between anaphoric and non-anaphoric uses of a given term, estimate the number of anaphors appearing in bibliographic records, and assess the effect on retrieval performance when anaphors are replaced by their referents. In the first phase of the study, rules were developed to form the basis for an automatic procedure to recognize anaphoric terms in bibliographic databases. An examination of the titles and abstracts of 600 documents revealed that only 3.67 true anaphors occurred in the average abstract, suggesting that the effect of treating these terms in some way to improve retrieval performance might be slight. In the second phase, 12 term weighting schemes were used to determine the relevance of each document to the corresponding query, and user's relevance judgements for the same searches were compared with the system's judgements for (1) searches using abstracts in which anaphors had been replaced with their referents, and (2) searches using abstracts with unresolved anaphors. These comparisons yielded mixed results, indicating that a straightforward substitution of referents for their anaphors will not improve retrieval performance in the majority of cases. It is concluded that future studies which treat document length more explicitly and study documents on an individual level are necessary. A bibliography is provided, and five lengthy appendices include the preliminary test and functional indexes, the retrieval experiment and functional indexes, results of the linguistic analysis, test results of rule sets, retrieval test results, and summaries of statistical results for searches of INSPEC and PsycINFO. (KM)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Research Supported By:
National Science Foundation
Grant NSF-IST-83/13716

ED277399

IMPACT OF ANAPHORIC RESOLUTION IN INFORMATION RETRIEVAL (INFORMATION SCIENCE)

Final Report July 1986

School of Information Studies
Syracuse University
Syracuse, New York 13244-2340

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jeffrey Katzer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR051742

BEST COPY AVAILABLE

IMPACT OF ANAPHORIC RESOLUTION IN INFORMATION RETRIEVAL

Final Report
October, 1986

Jeffrey Katzer
Susan Bonzi
Elizabeth Liddy

This material is based upon research supported by the National Science Foundation, Division of Information Science and Technology under Grant NSF-IST-8313716. The opinions, findings, conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation.

School of Information Studies
Syracuse University
Syracuse, New York 13244

This report was written by

Jeffrey Katzer, Elizabeth Liddy and Susan Bonzi

PROJECT STAFF

Principal Investigator	Jeffrey Katzer
Faculty Associate	Susan Bonzi
Research Associate	Elizabeth Liddy
Graduate Assistants	Elizabeth Oddy Joseph Janes
Primary Consultant	Robert N. Oddy
Project Secretary	Margaret Montgomery
Additional Assistants	Laurie Hussey Beth Livingston Richard Miller Lesley Pease Keith Williams

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the National Science Foundation, Syracuse University, and the School of Information Studies, as well as the approval and considerate involvement of both INSPEC and PsycINFO.

We also extend our appreciation to several individuals whose help contributed substantially to the project -- Elizabeth Oddy, for her control of the numerous computerized files, test of rule sets and her insights into various aspects of the project; Joseph Janes for writing a variety of computer programs for manipulation and analysis; Laurie Hussey, for testing and analysis of rules; Beth Livingston for editing of computer input; Richard Miller and Lesley Pease, for computer input; and Keith Williams for testing of rule sets.

We are particularly grateful to Professor Robert N. Oddy for his expertise, his vital assistance on very short notice, and his sincere concern for the outcome of the project.

ABSTRACT

Anaphora is the linguistic device of abbreviated subsequent reference to a concept. This research project was based on the hypothesis that within document frequency (WDF) of a term, and ultimately retrieval performance of a system using WDF, would be affected by the resolution of anaphora (replacement of its anaphor with its referent) within its text of a document. In order to test the hypothesis, a two-phase investigation was implemented.

In the first phase, all potential anaphors in a random sample of 300 abstracts from each of two databases were identified. Each occurrence of anaphora was then examined in order to determine if the term actually functioned anaphorically. From these observations, patterns emerged which were then developed into rules that captured the systematic regularities of functional anaphors. The rules were tested by at least three people to determine whether the rules accurately distinguished functioning anaphors from potential anaphors.

In the second phase of the project, 24 queries, abstracts retrieved from computerized searches on the queries, and relevance judgments on each retrieved document were selected from a previous research project. All functioning anaphors within the abstracts were resolved by hand. Twelve term weighting schemes were used on the basis of determining relevance of each document to its corresponding query. Two statistical relationships were then compared: 1) between the user's relevance judgment and the system's judgment based on the unresolved abstracts, and 2) between the user's relevance judgment and the system's judgment based on the resolved abstracts. If the latter relation is stronger than the former, then a formal treatment of anaphora in bibliographic retrieval positively affects system performance.

Results of the comparisons were mixed. In some instances, the resolved documents produced a significantly better correlation between user's judgments and system's judgments, while in other instances, the opposite occurred. The findings that resolution of anaphora may increase the performance of a retrieval are far from conclusive. It is clear that future studies of anaphora in information retrieval must be treated in a more complex manner than was attempted here.

OVERVIEW

In free-text information retrieval (IR) systems, all non-trivial words in the document are used to represent the content of that document. In the design of these systems, it is reasonable to believe that the more often a term is repeated, the more likely it is that the term represents a major concept of the document. It is for this reason that IR systems weight the importance of a given term as a function of its frequency of occurrence within the document. However, a straightforward count of each word type does not go far enough because it excludes ways in which the same concept can be represented by other words. Removing suffixes and combining synonyms are two methods that are used to make the resulting term weights better reflect the true presence of a concept in a document.

Another way in which an instance of a concept can be "hidden" in a count of term frequencies is through anaphoric reference, where, for example, a pronoun represents a major concept discussed elsewhere in the document. Though anaphora has been mentioned by several researchers in information science, very little is known about the extent of anaphora in bibliographic databases or how an explicit treatment of anaphora may change term weights and consequently retrieval performance.

This report documents the first investigation of anaphora in IR. More specifically, our objectives were to:

1. Develop procedures to recognize anaphors in text and to distinguish between anaphoric and non-anaphoric uses of a given term.

2. Estimate the number of anaphors appearing in bibliographic records.

3. Assess the effect on retrieval performance when anaphors are replaced by their referents.

These objectives are addressed in the next two sections of this report.

The first section is based on an examination of existing linguistic theory combined with a detailed study of a random sample of 600 documents (titles and abstracts). 142 words were identified as potential anaphors, though among the documents studied only 95 of these actually were present. These words were organized into ten classes and for each, rules were developed to determine whether a given term functioned anaphorically as it was used in the document. These rules can form the basis for an automatic procedure to recognize anaphoric terms in bibliographic databases. An examination of the 600 documents discovered that only 3.67 true anaphors occurred in the average abstract -- suggesting that the effect of treating these terms in some way to improve retrieval performance might be slight.

The second section of this report presents the results of our examination of the third objective. This study is based on the premise that anaphors are used by authors to avoid repetition and as such, they are likely to represent the more important concepts in a document. Therefore, replacing all anaphors with their referents will change term frequencies in such a way so as to improve retrieval performance. A post-retrieval experiment was conducted making use of 12 existing queries for each of the two bibliographic databases. All documents retrieved by these queries were examined to identify all true anaphors. Then, by hand, each of these anaphors was replaced with its word or phrase referent. This process changed the frequency of occurrence of words in the document, and therefore the predicted relevance of the retrieved documents was also changed. If the process of replacing anaphors with their referents improves retrieval performance, then the revised set of term frequencies should predict document relevance better than the original frequencies.

The results of the study are mixed. Treating anaphora does improve retrieval for several queries though all classes of anaphora do not contribute equally to this improvement. There are also instances in which retrieval performance decreases when given classes of anaphora are replaced with their referents. However, for the majority of queries there is no effect of treating anaphora in this way. The major conclusion of this work is that a straightforward substitution of anaphors for their referents will not improve retrieval performance in the majority of cases. We remain convinced, however, that the basic premise underlying this research is true, viz., that anaphors are used to abbreviate subsequent mentions of the more important concepts in a document. Therefore, the study of anaphora in IR research should not be abandoned, rather, other means of isolating the reference to key concepts need to be explored.

Two avenues of additional work are proposed. First, document length needs to be treated more explicitly. When an anaphor is replaced it often is not a one-word for one-word substitution. Instead, entire phrases may be added to the document, increasing the number of trivial terms more than the number of instances of key terms. Because ranking formulas tend to be sensitive to the total number of words in a document, retrieval performance can deteriorate after an anaphor is replaced. Another approach to limiting the increase in document length is to edit the substitution process by allowing only terms that appeared in the query to be added to the document when an anaphor is replaced. The second area of additional work is to study documents on an individual level. By focusing on retrieval performance, our level of analysis had to be the query. Replacing anaphors with their referents may affect individual documents quite differently and the overall effect on the query would be some "average" of what happened to the individual documents.

At this time, these two areas of future work seem to hold the most immediate promise for tapping the potential of using the full semantic content of anaphors to improve information retrieval effectiveness. This potential effect exists not only for document abstracts used in free-text searching, but also in other areas of information retrieval work that use naturally occurring texts such as users' queries or full-text documents in a question-answering system.

TABLE OF CONTENTS

Page

I. OVERVIEW i

II. A STUDY OF DISCOURSE ANAPHORA IN SCIENTIFIC ABSTRACTS 1

- A. Introduction. 1
- B. Study 7
- C. Results 12
- D. Discussion. 16
- E. References. 18

III. THE EFFECTS OF ANAPHORIC RESOLUTION ON RETRIEVAL PERFORMANCE 20

- A. Introduction. 20
- B. Anaphora. 21
- C. Method. 24
- D. Results 30
- E. Discussion. 35
- F. References. 38

BIBLIOGRAPHY 40

APPENDICES 44

- A. Preliminary Test, Functional Indexes. 45
- B. Retrieval Experiment, Functional Indexes 55
- C. Linguistic Analysis 67
- D. Test Results of Rule Sets 131
- E. Retrieval Test Results. 134
- F. Summaries of Statistical Results INSPEC and PsycINFO 354

LIST OF TABLES

Page

"A Study of Discourse Anaphora in Scientific Abstracts"

Table 1, Distribution Across 2 Subject Domains 13

Table 2, Class Summary 15

Table 3, Testing of Rules 17

"The Effects of Anaphoric Resolution on Retrieval Performance"

Table 1, Classes of Anaphora 23

Table 2, Database Population and Samples 26

Table 3, Term-Weighting Schemes 28

Table 4, Analysis 31

Table 5, Summary of Statistical Results 34

A Study of Discourse Anaphora in Scientific Abstracts[1]

Elizabeth Liddy, Susan Bonzi, Jeffrey Katzer & Elizabeth Oddy
School of Information Studies,
Syracuse University, Syracuse, New York 13244

Introduction

Much of the work that information retrieval is involved in makes use of naturally occurring texts such as users' queries, abstracts in a free-text retrieval system, or full-text documents in a question-answering system. To develop successful systems in any of these areas requires an adequate handling of the whole range of linguistic phenomena that exhibit themselves in naturally occurring text. They may be word-level (morphology) or sentence-level (syntax) phenomena or they may be discourse level phenomena which become a factor when analyzing units of text larger than a single sentence. Designers of information retrieval systems have already learned to apply linguistic knowledge developed in both morphology and syntax. For example, morphology has contributed the technique of stemming which conflates terminological variants to their stem, while the automatic identification of noun phrases for use as indexing phrases uses syntactic analysis [1]. However, information retrieval systems which manipulate chunks of connected text must also attend to the text level phenomena which have more recently come under study in discourse linguistics. Among the linguistic devices of concern at the discourse level are anaphora, cataphora, ellipsis, substitution, parallelism and inter-sentential conjunction.

---

[1] Based on an article accepted for publication in JASIS.

The discourse level phenomenon dealt with in this paper can be most inclusively referred to as discourse anaphora. This use of the term anaphora reflects common usage in discourse linguistics rather than that of Chomsky and linguists of the transformational grammar school who use the term 'anaphor' in a more narrowly defined sense. While Chomsky is concerned with determining the exact conditions under which pronouns function within one sentence, our concern is with all anaphoric-type references, whether within or across sentence boundaries. Discourse anaphora can be defined as abbreviated subsequent reference and is most commonly exemplified by, but not limited to, the use of pronouns. Examples of discourse anaphora can be seen in the following excerpt where the term "counteridentification" is actually used only once, but the concept is semantically present a total of three times since it is anaphorically referred to twice more, once by "this mechanism" and once by "it".

> Counteridentification is a mechanism that makes changes within the psychic structure of the individual. This mechanism differs from negative identification in that it uses the aggressive energies....

Humans (e.g. indexers, or users judging document relevance) mentally resolve anaphoric references and appear able to take abbreviated references into consideration in constructing appropriate mental representations of text. This is facilitated by the fact that in expository texts a new entity (a concept or object) is usually introduced to the reader in its fullest, most explicated form. A possible syntax for such a noun phrase is:

det + adj_1 + ... adj_n + noun + prep phrase/rel clause

Full first-mention is used in order to firmly establish a virtual instance of the entity in the mind of the reader. Having successfully anchored the lexical realization to a mental representation, further comments can be made about that entity without repeating all the pre- and post-modifiers used in the first-mention realization form, or even without using the noun itself.

The range of possible subsequent-mention realization forms which would be considered anaphoric references, include:

- determiner + same noun
- determiner + general noun
- pronoun

All of these subsequent-mention forms are shorter and convey less information than full first-mentions. However, these forms do communicate successfully and unambiguously to a reader because all the text need do is remind the reader which entity is being mentioned, rather than create a new mental representation.

Although humans seldom encounter difficulty in recognizing an anaphor and correctly identifying the referent of the anaphoric expression in text, discourse anaphora remains one of the text level phenomena still posing substantial difficulties for the many fields that are attempting to make use of naturally occurring texts. In information science, the necessity for recognizing and resolving anaphoric references impacts on 1) natural language understanding, 2) question-answering, 3) automatic extracting, 4) query analysis, and 5) bibliographic retrieval.

Natural Language Understanding: A natural language understanding system needs to build a semantic representation of the text being processed. In order to do this successfully, the difficult task is not in accurately representing the meaning of each new input sentence singly, but rather in appropriately combining the meaning of all individual sentences to form a representation of the aggregated meaning of the text. It is a matter of interpreting new information in light of the old and of connecting new information to the appropriate old information in the representation, so that a coherent whole results. It is not unexpected, then, that the task of correctly interpreting discourse anaphora is essential for building integrated representations of meaning for natural language understanding systems [2].

Question-Answering: Question-answering systems may be of two types, both of which require handling of discourse anaphora. One approach to question-answering systems is to build semantic representations of both the texts in the system and the users' queries and use the latter representations to find appropriate answers among the former. If this is the approach taken, the rationale given above for anaphora resolution techniques in N.L.U. holds for this task as well. An alternative approach to question-answering has been attempted by John O'Connor [3]. O'Connor attempted to provide answers to queries by retrieving answer-providing passages from the actual text of the document rather than building an intermediate semantic representation of the text. His results were very promising, but O'Connor suggested that further improvement could be gained if it were possible to locate in text the fully explicated expressions which are subsequently referred to in an abbreviated manner by anaphoric clues such as 'this', 'these' and 'those'.

Automatic Extracting: Paice's work [4] on automatic extracting clearly recognized the need for attending to anaphoric reference in text. In order to automatically compile a comprehensible, substantive extract, Paice found it necessary to establish a list of 'clue words' (e.g. 'it', 'them', 'similar', 'both') which indicated that if the particular sentence in which these words occurred was to be included in the extract, it would be necessary to locate and include the earlier text in which these anaphoric references were more fully explicated.

Query Analysis: Research currently underway by Oddy [5] (see also Belkin, Oddy, Brooks [6]) into an information seeker's state of knowledge on the topic or problem which compelled their interaction with the information retrieval system, includes techniques for analyzing and representing relationships between concepts in the user's problem statement. These relationships are currently computed from quite superficial, mainly statistical, characteristics of the texts. Also, the texts are transcripts of oral utterances with copious use of anaphora. Hence, resolution of discourse anaphora would undoubtedly affect the derived representation of the user's state of knowledge.

Bibliographic Retrieval: In free-text document retrieval systems, the problem of correctly recognizing and resolving subsequent references is important because many of the statistical methods of determining which documents are to be retrieved in response to a query make use of frequency counts of terms. For this count to be a true measure of semantic frequencies, it would appear that the semantically reduced subsequent references should be resolved by their earlier, more fully specified referents in text. A technique in many experimental and a few operational document retrieval systems is to weight the terms of a document's free-text representation (title and abstract) on the basis of term frequencies. The information retrieval system, applying a similarity measure between query and document representations, will then do a best-match search, retrieval, and ranking of documents for the user. This technique is based on two apparent assumptions: 1) that frequency of occurrence is a good indication of the degree to which a piece of text is about a certain term, and 2) that an adequate means of determining semantic frequency of a concept is by counting all explicit occurrences of a term.

However, the theory behind discourse anaphora predicts that an adequate measure of frequency of occurrence of a concept requires that all implicit occurrences of that term be taken into account. In bibliographic retrieval, this would mean that the frequency count of document terms after resolution of all anaphors would better represent what the document is about and that resolving anaphoric terms in abstracts would significantly improve retrieval results by obtaining for the user a ranked ordering of documents more truly reflective of the documents' degree of relevance to the user's query.

Even though the areas of work in information science discussed above need to be concerned with discourse anaphora, no study exists which provides either 1) base-line quantitative data on the extent to which this phenomenon exists in a text-type used in information science, or 2) insight into whether the use of anaphoric references in such a text type is rule-governed enough to permit development of algorithms for automatically detecting and then resolving anaphoric references. In fact, a recently published investigation by Fidel [7] of those aspects of free-text which might impact on retrieval, mentioned no concern about anaphoric references. It is hoped that the benchmark descriptive data and feasibility of automatic recognition and resolution of anaphora provided by this study may be useful to those areas of work on which the presence of anaphoric terms has an impact.

Study

Our work consisted of detecting occurrences of anaphoric references and computing base-line counts, as well as developing rules which would capture in algorithmic form the decisions made by human processors both as to whether a term is anaphoric or not and to what is its proper referrent. Before these tasks could be attempted, however, some preliminary steps were required.

The first task was to develop a list of all those terms considered potentially anaphoric. Having located no such pre-established, all-inclusive list in the literature, we compiled a list from grammar books, particularly Quirk, Greenbaum, Leech & Svartik [8], linguistic works dealing with linguistic devices adding to the cohesion of a text, such as Halliday & Hasan [9], and Grimes [10], and prior investigations into some subset of the phenomena of discourse anaphora (Webber, [11], Sidner, [12], and Hirst, [2]). This resulted in the set of 142 potential anaphors (P.A.s), listed in (Figure 1).

*above
*additional
aforementioned
aforesaid
*all
*another
another's
*any
anybody
*anyone
*anything
*both
*did
*do
*does
*doing
*done
*each
eighth
*either
*else
else's
elses'
*enough
*equal
*every
everybody
everyone
everyone's
everything
*few
*fewer
fewest
*fifth
*first
forementioned
*former
former's
*fourth
*he
*her
*here
hers
*herself
*him
*himself
*his
*I

*identical
*identically
*it
*its
*itself
*last
*latter
latter's
*least
*less
likewise
*little
*many
me
mine
*more
*most
*much
my
myself
*neither
*ninth
*no
nobody
*none
*nothing
*one
*one's
ones
*ones'
*other
other's
*others
*our
ours
ourselves
*S
*S's
*Ss
*Ss'
*same
*second
seventh
*several
*she
*similarly
*sixth
*so

*some
somebody
somebody's
someone
someone's
*something
*such
tenth
*that
*the
*their
theirs
*them
*themselves
*then
*there
thereat
therefor
therefrom
*therein
thereinto
thereof
thereon
thereout
thereto
thereunder
therewith
*these
*they
*third
*this
*those
*thus
*us
*vice versa
*we
*where
*which
*who
*whom
*whose
you
*your
yours
yourself
yourselves

Figure 1

A classification scheme (Figure 2) was then imposed on these

P.A.s so that work could proceed at the class or sub-class level. The class distinctions were made on a functional basis guided by the intuition gained from a small feasibility study which investigated whether recognition and resolution techniques would be generalizable at the functional class level.

1. Central Pronouns
    a. Personal Pronouns - he, him, it
    b. Possessive Pronouns - his, her, their
    c. Reflexive Pronouns - itself, themselves
2. Nominal Demonstratives - this, these, those
3. Relative Pronouns - who, which, where
4. Nominal Substitutes - above, former, one
5. Pro-verb - do
6. Indefinite Pronouns - any, each, many
7. Pro-adjectives - another, identical
8. Pro-adverbials - so, such, similarly
9. Subject References - S, Ss
10. Definite Article - the

Figure 2: Classes of Discourse Anaphora with examples

Most of the terms which are capable of anaphoric reference can also perform other functions in text and as a result should be considered as only potential anaphors (P.A.s). Any system which adequately handles subsequent reference in text first needs a means for determining in a particular instance if a P.A. is actually a functioning anaphor (F.A.). Although most humans can quite easily decide in a specific instance whether a term is being used anaphorically or not, the precise linguistic evidence on which these decisions are made is not available in the literature and needs to be delineated, so that algorithms can be written for accomplishing the same task. Therefore, we conducted a study to see whether it would be possible to develop rules which could be successfully applied by independent judges and result in a clear separation between those instances where a P.A. is simply a P.A. and those instances where a P.A. is an F.A. Success with these rules would suggest the feasibility of developing machine-implementable algorithms to make the same distinctions.

To write such algorithms, it is necessary to look at a corpus sufficiently large that the regularities of syntax and lexical choice which would serve as the basis of these rule-based algorithms will exhibit themselves. Unfortunately, much of the previous work in linguistics on anaphora has used contrived texts or corpuses too small to generalize from. So, for a corpus on which to write and test rules for recognizing when a P.A. is an F.A., we drew 600 abstracts at random, 300 each from two operational document retrieval databases: 1) PsycINFO - which contains abstracts of documents reporting on the behavioral sciences, and 2) INSPEC - which contains abstracts of documents reporting on engineering and computer science. This combined set contained occurrences of 95 P.A.s (starred terms in Figure 1) from the preliminary compilation of 142 P.A.s. These 95 terms, on which the following work is based, were assigned to one of the 10 classes of anaphoric terms (see Figure 2).

The basic procedure which was followed in developing and testing the P.A.-to-F.A. rules is outlined as follows:

1. For each class, all abstracts containing occurrences of terms of that class were collected. The exact number of abstracts drawn for each class correlated roughly with the frequency with which terms of that class occurred.
2. For each occurrence of a P.A., an intellectual decision was made as to whether the P.A. was an F.A.. This provided the basic summary data being reported here.
3. While doing the above step, patterns began to emerge from the texts: the predictability of contextual information in determining whether the use of the term was anaphoric or nonanaphoric became evident.
4. From these observations, P.A.-to-F.A. rules were written which capture the systematic regularities which, when encoded in algorithms, will, we hope, replace human intuitive decision making. These regularities are either in the lexical environment in which anaphoric/nonanaphoric use of a term can be predicted to occur, or the particular syntactic construction indicating anaphoric /nonanaphoric use.[2]
5. The P.A.-to-F.A. rule sets for each class, sub-class, or term were slightly reworded where necessary using a less linguistically oriented vocabulary. Each rule set was given to at least three judges who applied them to a subset of the original 600 abstracts. Each rule was tested on ten

---

[2] Rules have not as yet been developed for class 10, the definite article, due to the unpredictability of the contexts in which 'the' appears. Following analysis of the results of the retrieval experiment, rules will be attempted if the results warrant algorithm development for this class.

different occurrences of the term(s) to which the rule applied. If there were less than 10 occcurrences, all of the available occurrences were tested.

Results

The results reported here are of a twofold nature: 1) summary data on distribution of P.A.s and F.A.s in abstracts; and 2) success of writing rules for use in determining whether a P.A. is an F.A.

Distributional Analysis

The summary data indicates that the linguistic phenomenon of discourse anaphora exhibits itself to a greater extent in PsycINFO than in INSPEC.

Table 1 shows the mean occurrence of P.A.s per abstract to be 13.2 for the PsycINFO abstracts, and 10.08 for the INSPEC abstracts, with a mean occurrence of 11.64 P.A.s per abstract across the complete sample of 600 abstracts. The mean occurrence of F.A.s per abstract is 4.49 for the PsycINFO abstracts, and 2.86 for the INSPEC abstracts, with a mean occurrence of 3.67 F.A.s per abstract across the complete sample of 600 abstracts. These preliminary results suggest that the phenomenon of discourse anaphora has a greater impact on a natural language text-handling systems in the behavioral sciences as compared to computer science and engineering.

These results might appear to suggest that since there are far fewer F.A.s than P.A.s, the effects of resolving F.A.s may not be as large as a casual study of P.A.s would indicate. It should be noted, however, that since discourse anaphors are used by the writer to avoid needless repetition, anaphors are more likely to be used to replace the major concepts in a piece of text. As a result, resolving even the mean of 3.67 F.A.s per abstract may have a strongly differential impact on term frequencies and ultimately, retrieval results. Also, since most pieces of text are organized around one or two major concepts, the effect of leaving anaphoric references untreated has the same potential for substantive impact on any of the information science areas which deal with naturally occurring texts.

Table 1: Distribution across 2 Subject Domains

	P.A.s		F.A.s	
	No.	Mean	No.	Mean
PsycINFO	3960	13.2	1347	4.49
INSPEC	3024	10.08	857	2.86
Total	6984	11.64	2204	3.67

A Functional Index (F.I.=#F.A/#P.A.) was computed for each class in each database (Table 2). Appendix A contains this information for each individual term in the set of 600 abstracts and Appendix B, figures for the 487 documents used in the retrieval experiment. The F.I. is an important parameter of consideration as we are interested in developing resolution algorithms for those classes in which a high proportion of the

P.A.s are F.A.s. Given only the information in Table 2, some classes appear far likelier candidates than others. For example, of the 493 uses of central pronouns, 78% of the occurrences were anaphoric. This high F.I. contrasts with the use of the definite article 'the' which has a very high frequency of occurrence (3435 uses across both databases) yet an F.I. of only 14%. On this basis, it would be unlikely that one would choose to devote one's efforts to developing algorithms for classes with so low an F.I.. Yet the results of the retrieval experiment will also be taken into consideration when choosing classes for algorithm development. It makes sense to concentrate our efforts on those classes with both a high F.I. and demonstrated positive effect on retrieval performance.

Rule-Governed Recognition of Functional Anaphors

The other area of results to be reported is that of the extent to which the rules for deciding when a P.A. is an F.A. can successfully be applied by independent judges. These results provide preliminary evidence of whether the environment in which an anaphoric usage occurs is predictable enough to make automatic recognition possible. Three judges were used for testing each set of rules. The judges were not aware that their decisions were on the anaphoricity of a term. They were instructed to follow a set of rules which described distinct patterns of usage of a term and decide which pattern a particular instance matched. The rules used by the judges were based on the linguistic regularities observed and captured in individual analyses of each functioning anaphoric term and are contained in Appendix C.

Table 2: Class Summary

	PsycINFO			INSPEC			Totals		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
Central Pronouns	244	60	.80	143	46	.76	387	106	.78
Nominal Demonstratives	176	265	.40	155	148	.51	331	413	.44
Relative Pronouns	227	255	.47	192	88	.69	419	343	.55
Nominal Substitutes	60	63	.49	64	71	.47	124	134	.48
Pro-verb	21	42	.33	3	12	.20	24	54	.31
Indefinites	128	317	.29	44	209	.17	172	526	.25
Pro-adjectives	27	51	.35	10	40	.20	37	91	.29
Pro-adverbials	25	52	.32	30	68	.31	55	120	.31
S & Ss	188	25	.88	0	0	-	188	25	.88
Definite Article	251	1483	.14	216	1485	.13	467	2968	.14
Totals	1347	2613	.34	857	2167	.28	2204	4780	.32

The rule sets consisted of an ordered series of pattern matching tasks against either syntactic or lexical templates. Judges decide whether a usage matches Rule 1, or Rule 2, and so on, down the list of rules for that class or term. Some of these rules define anaphoric uses, others nonanaphoric, but the judge was not concerned with this. The judges' decisions were strictly governed by the pattern matching aspect of the rules. These are the types of human decisions that algorithms are able to mimic and which will make the automatic recognition of anaphoric uses of terms possible. The eventual automation of this task would require, in addition to the algorithms, the inclusion of two components commonly avaliable in text processing systems, namely, a parser and a lexicon with semantic class information. (See [13] for a sample set of rules for the Nominal Demonstrative 'that'.)

Table 3 presents the average rate for classes 1-9, across three judges, of correctly applying the rules for deciding which pattern of usage a particular instance of a term follows. Appendix D contains the success rate of applying rules for each judge and cumulatively for each term tested. The success of the pattern-matching rules in correctly predicting the same decision as an overt intellectual decision on a term's anaphoricity ranged from a low of 83% for the terms comprising the S & Ss class to a high of 99% for the proverb 'do'. These initial results give us confidence in the field's ability to develop P.A.-to-F.A. algorithms, particularly since an error analysis has identified the recurring problem with the rules to be a difficulty in deciding when a subsequent definite noun phrase containing a class level noun refers to the same entity as a previous specific noun (e.g. 'the instrument' uses the P.A. 'the' plus a class level noun used as a less specified reference to a particular test instrument mentioned earlier in text). Inclusion of semantic class information in the system's lexicon could easily lessen the number of errors of this sort.

Discussion

With a mean occurrence of 3.67 functioning anaphors per abstract across the full sample of 600 abstracts, this study indicates that terms capable of anaphoric reference occur suffi-

Table 3: Testing of Rules

1. Central Pronouns	98%
2. Nominal Demonstratives	87%
3. Relative Pronouns	93%
4. Nominal Substitutes	88%
5. Pro-Verb	99%
6. Indefinites	89%
7. Pro-adjectives	86%
8. Pro-adverbials	96%
9. S & Ss	83%

ciently frequently in abstracts, to raise questions as to the adequacy of techniques which use surface counts of a term as a sufficient measure of the total times that a concept is referred to in an abstract. In that anaphors tend to be used for shortening the reference to the major concepts of a text, it is intuitively clear, although awaiting empirical proof, that resolution of these anaphoric references will generate term frequencies which provide better representations of the information content of documents and improve retrieval in an operational setting. These representations will be based on the frequency of reference to a *concept* rather than the currently used frequency of occurrence of a *term.*

In the second phase of this research project, we conducted an experiment on the impact of resolving anaphors in one area of information science, namely bibliographic retrieval, but consider the effect to be more far ranging than just retrieval, especially since the numbers of F.A.s is a function of the length of the text. Any work with naturally occurring text is affected by the linguistic phenomenon of discourse anaphora. As noted above, work in the areas of question-answering, automatic extracting, and query analysis have acknowledged the need to develop techniques for handling anaphoric terms. We are hopeful that our results will provide some previously unavailable base-line data on discourse anaphora in one particular text-type across two subject domains.

Results of the rule testing indicate that algorithms for determining automatically whether a potentially anaphoric term is functioning as an anaphor in a particular instance are indeed feasible since the task has been shown to be one of pattern matching governed by rules applied with high reliability. In addition, a similar algorithmic approach for resolving functioning anaphors with their appropriate referrents will be suggested for several of the classes of anaphors after a full analysis of the retrieval experiment results is completed.

References

1. Waldstein, R. The Role of Noun Phrases as Content Indicators. Ph.D. Dissertation, Syracuse University School of Information Studies, 1981.

2. Hirst, G. Anaphora in Natural Language Understanding: A Survey. New York: Springer-Verlag; 1981.

3. O'Connor, J. "Text Searching Retrieval of Answer-Sentences and Other Answer-Passages." Journal of the American Society for Information Science. 24(6): 445-460; 1973.

4. Paice, C. D. "The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self-Indicating Phrases." In: Oddy, R.N., ed. Information Retrieval Research. London: Butterworths; 1981: 172-191.

5. Oddy, R.N. A Study of Representations for Anomalous States of Knowledge in Information Retrieval. NSF Grant Proposal IST-8420608. Syracuse, New York: Syracuse University School of Information Studies; 1984.

6. Belkin, N.J.; Oddy, R.N.; Brooks, H.M. "ASK for Information Retrieval: Part I. Background and Theory. Part II. Results of a Design Study." Journal of Documentation. 38 (2,3): 61-71, 145-164; 1982.

7. Fidel, R. "Writing Abstracts for Free-Text Searching." Journal of the American Society for Information Science. 42(1): 11-21; 1986.

8. Quirk, R.; Greenbaum, S.; Leech, G.; Svartvik, J. A Grammar of Contemporary English. London: Longmans; 1972.

9. Halliday, M.A.K.; Hasan, R. Cohesion in English. London: Longmans; 1976.

10. Grimes, J. The Thread of Discourse. The Hague: Mouton Publishers; 1975.

11. Webber, B.L. A Formal Approach to Discourse Anaphora. New York: Garland Publishing, Inc.; 1979.

12. Sidner, C. Towards a Computational Theory of Definite Anaphora Comprehension in English Disourse. Artificial Intelligence Laboratory Report TR-537; Cambridge, Massachusetts: Massachusetts Institute of Technology; 1979.

13. Liddy, E.; Bonzi, S.; Katzer, J.; Oddy, E. "A Study of Discourse Anaphora in Scientific Abstracts." Journal of the American Society for Information Science. (In press).

THE EFFECTS OF ANAPHORIC RESOLUTION ON RETRIEVAL PERFORMANCE [1]

Jeffrey Katzer, Susan Bonzi, Elizabeth Liddy

Syracuse University, School of Information Studies
Syracuse, New York 13244

INTRODUCTION

For almost thirty years, work in automatic indexing has been a major component of information retrieval research. To perform effectively, indexing schemes must be able to accurately portray what the document is about and they must assist in the discrimination among documents in the collection. Within document frequency (WDF) is clearly helpful in meeting the first of these functions: the more often a term is used in a document, the greater the likelihood that the concept or subject underlying the term is central to the document. However, it is not clear that WDF is of much value to the second function of indexing schemes. For a term to distinguish among documents, its WDF must have a large variance over the collection. Given that documents are composed of relatively few words, such as titles and abstracts, coupled with the rather mechanical means for automatically recognizing a given term (e.g. counts of synonyms are usually not combined), it is doubtful that any sizable variance in WDF would occur. Some support for this contention is provided by Sparck Jones [2] who found that 62% of the terms in one small database had a WDF of one and another 19% of the terms had a WDF of two.

There is little evidence of the effect of increasing the variation of WDF on retrieval performance. It is not enough to simply increase the WDF weights. To be of value in retrieval the increase must be disproportional, raising the variability by affecting key terms more than other terms. Using longer documents (full text instead of abstracts) is one method to accomplish this. Another is to bring together into one class, all mentions of a single concept -- whether referred to by the same root, by a synonym, or by the linguistic technique known as anaphora. Stemming of suffixes is common to most approaches to automatic indexing and thesauri have been used to combine synonyms into a single class. However, the effect of anaphora on WDF and ultimately on retrieval performance has not been studied.

ANAPHORA

Anaphora, briefly defined, is the linguistic device of abbreviated subsequent reference. Consider the following sentence [3]:

> Wash and core six baking apples and place them in a pan.

The pronoun, "them" is an anaphor and is easily understood by people to mean, "six washed and cored baking apples".

---

1. This report is based on a paper presented at the 1986 ASIS Annual Meeting, Chicago, Illinois.

Anaphora is one of several so-called cohesion devices used in both written and spoken discourse to (1) avoid monotonous repetition, (2) shorten the discourse, and (3) enhance the coherence of the passage. Because anaphors are used to eliminate repetitiousness, they are more likely to be used to replace the major concepts and terms in an abstract. Thus, we would expect that resolving all anaphors in an abstract will increase the WDF of important terms proportionally more than it will raise the frequencies of other terms.

Although human intellect has no difficulty recognizing and resolving anaphors (replacing them with their referents), automatic methods to accomplish these tasks are still in their infancy. Work in natural language understanding has made some advances in the treatment of anaphora, but that work is restricted to limited subject domains or certain classes of anaphora. [4-10] In information science, anaphora is almost completely ignored. There is some mention of it in the literature [11-14], but not in terms of document retrieval. Instead, anaphora is considered in the treatment of question-answering systems, passage retrieval, or automatic abstracting.

Table 1 presents the major classes of anaphora used in the current study; see [15] for a more complete description. Whether or not a given anaphor actually functions anaphorically can only be determined by analyzing the linguistic context within which the term exists. Thus, the

TABLE 1

CLASSES OF ANAPHORA

	ANAPHORIC CLASS	EXAMPLES
A:	CENTRAL PRONOUNS	they, their, themselves
B:	NOMINAL DEMONSTRATIVES	this, that, these, those
C:	RELATIVE PRONOUNS	who, which, where
D:	NOMINAL SUBSTITUTES	above, former, one
E:	PRO-VERBS	do
F:	INDEFINITES	some, all, each
G:	ADJECTIVES	another, both, identical
H:	ADVERBS	so, such, similarly
I:	SS	(subjects)
J:	DEFINITE ARTICLE	the

resolution task depends upon (1) an exhaustive list of all potential anaphors, (2) a set of rules to determine if a particular potential anaphor is actually functioning anaphorically, and (3) a set of rules for replacing the functioning anaphors with their referents.

METHOD

In this study, retrieval performance depends upon the degree to which the predicted relevance of "unresolved" and "resolved" documents matches the user's relevance judgments. Thus, three sets of judgments are needed: (1) those based on the user's assessment of documents retrieved by an IR system in response to a query, (2) those produced by the retrieval system from unresolved stems in the document, and (3) those produced by the system from the resolved stems in the document.

Databases, Queries, & Relevance Judgments: Since the three relevance judgments noted above can be produced in a post-retrieval experiment, queries and relevance judgments collected in other studies could be re-analyzed for our current work on anaphora. Only a brief description of these existing materials will be provided here; a fuller accounting can be found elsewhere. [16]

Two databases were used to increase the generality of the findings. Each was composed of approximately 12,000 documents consisting of a citation and an abstract of 75-175 English words. From the earlier study, we had 84 queries to the INSPEC database and 57 queries to PsycINFO. All queries were posed by individuals with genuine information needs and were searched by trained intermediaries. The relevance of the retrieved documents was determined by the originator of the query using a four-point categorical scale: 1 being highly relevant, 2 slightly relevant, 3 slightly non-relevant, and 4 highly non-relevant.

The current research could make use of only a small subset of the available queries. Some queries had to be excluded because there was insufficient variability in the relevance judgments assigned to the retrieved documents. Others were excluded to decrease the amount of work involved in identifying and resolving "by hand" all anaphors in all retrieved documents. Queries were selected which met the following criteria: (a) the number of retrieved documents ranged between 15-30, (b) there were at least two retrieved documents judged at each of the four relevance categories, and (c) no more than 60% of the retrieved documents were judged relevant -- in categories 1 or 2. These criteria selected 12 queries from INSPEC and 17 from PsycINFO. Five queries were randomly discarded from PsycINFO to make the two sets equal. Table 2 describes these two test collections.

TABLE 2

DATABASE POPULATIONS AND SAMPLES

	INSPEC	PsycINFO
AVAILABLE DATA		
NUMBER OF DOCUMENTS	12,864	11,662
NUMBER OF USER QUERIES	84	52
NUMBER OF TYPES IN DATABASE	67,401	35,758
SAMPLE USED		
NUMBER OF QUERIES	12	12
DOCUMENTS RETRIEVED BY EACH QUERY	12	15-25
UNIQUE DOCUMENTS IN ALL QUERIES	261	226

Predicted Relevance from Resolved and Unresolved Documents: For each of the 487 retrieved documents all potential anaphors were identified by comparing each term in the file of documents with a "dictionary" of all anaphors that occurred in the two databases. Over 6500 potential anaphors were found. Each of them was inspected within its linguistic context to determine if it actually functioned anaphorically in the document; over 2200 true anaphors were identified. The final step was to apply another set of rules to resolve all functioning anaphors. The dictionary of anaphors and the rules to discriminate between potential and functional anaphors were developed and validated on other document samples from the two databases. [15] At this point, two collections of the 487 documents existed: one as originally contained in the database and one with all anaphors replaced with their referents.

Term-weighting Schemes and Similarity Measures: Different methods for weighting terms and for determining the degree of similarity between documents and the query affect retrieval performance differently. [17] Therefore, it was important to consider alternative approaches to term-weighting and similarity. Table 3 lists the 12 term-weighting schemes employed in the study. Most of these include WDF -- either alone, corrected for length, or in combination with collection frequencies. Collection frequencies and term postings were tabulated separately using slightly different methods of processing; for

TABLE 3

TERM-WEIGHTING SCHEMES

(a)	1	(g)	$f/\log(k)$
(b)	$1/t$	(h)	f/F
(c)	$1/\log(t)$	(j)	$f/\log(F)$
(d)	f	(l)	$f/[(k)(F)]$
(e	$\log(f)$	(m)	$f/[\log(k)(F)]$
(f)	$f/(k)$	(n)	$[f][\log(N/d)]$

d = number of postings of term;
f = within document frequency;
F = frequency of term in database;
k = number of tokens in document;
N = number of documents in database;
t = number of types in document.

approximately eight percent of the terms, the number of postings was higher than the collection frequency. These differences prevented our use of any term-weight that combined with postings and collection frequencies. WDF, however, was not affected by these differences. The cosine correlation and Dice's coefficient were the two similarity measures used. Combining the term weights with the similarity measures yielded 19 different pairs. *

Analysis: The major research question can be answered by comparing two statistical relationships: (1) that between the users' relevance judgments and system's relevance based on unresolved anaphora, and (2) that between the users' judgments and the system's relevance based on resolved anaphora. If the second relationship is stronger than the first, it may be reasonable to conclude that resolving anaphora in a document will affect WDFs in such a way so as to improve retrieval performance.

There are thousands of relationships to be compared. Each combination of term-weighting scheme and similarity measure was used separately on each of the ten classes of anaphora (see Table 1) and on an "eleventh" class, made up of the union of the other ten. This entire set of analyses was carried out for all queries in the two databases. Each relationship was quantified using Pearson's well-known

---

* With the Cosine Correlation, term weights b and c are equivalent to a; f and g are equivalent to d; and L is the same as H.

measure of linear correlation. The two correlations (between the users' judgments versus resolved documents and users' judgments versus unresolved documents) were compared to see if one is statistically higher than the other. The analysis plan is summarized in Table 4.

RESULTS

Clearly, with over 5000 combinations of results to consider, it is difficult to draw simple conclusions. Moreover, care must be taken in interpreting individual findings because statistically some 250 tests (of differences between the two correlations) could achieve significance at the .05 level by chance alone. Therefore, the general patterns of results shown in Table 5 and on Appendix F will be examined rather than the raw findings given in Appendix E.

In general, the results are mixed. For the majority of queries, replacing anaphors with their referents did not have any real (non-chance) effect on the predicted order of document relevance.

Some resolutions had a negative effect, i.e. resolving anaphors reduced the retrieval performance in terms of ranking. The most obvious example of this is INSPEC Query #109 which had negative results in four different classes of anaphora. The most likely explanation for negative findings may be document length. Resolving anaphors does not simply

TABLE 4

ANALYSIS

For a given query, TW, and SM, do

UNRESOLVED RANKINGS

1. Rank documents by predicted relevance
2. Correlate these with user's relevance

RESOLVED RANKINGS

3. Resolve a single class of anaphora in all documents
4. Recompute TWs and SMs
5. Rank documents by predicted relevance
6. Correlate these with user's relevance

COMPARE #2 vs. #6

7. Determine which set of rankings better match the user's judgments

REPEAT ALL OF THE ABOVE FOR

A. All combinations of TWs and SMs (19)
B. All classes of anaphora and a combined class (11)
C. All queries (12)
D. All databases (2)

replace a single work (such as a pronoun) with another simple word (such as the noun to which the pronoun refers). Instead, anaphors may need to be resolved with phrases of several words -- most of which can be trivial. Since some of the term-weighting schemes and the similarity measures were not corrected for document length, resolution could, in these cases, have had a negative effect.

However, it is also evident from Table 5 that resolution increased retrieval performance for several queries -- #158, #180, #203, and #212 seem most obvious. It is worth noting that positive effects for several anaphoric classes do not necessarily accumulate into an overall positive effect when all classes are resolved (Class R): for #158 there isn't any overall effect, while for #212 the overall effect is mixed. There is no clear pattern of what is required to obtain a positive result in Class R -- compare query #107 with #170, or #221 with #222. Obviously, total resolution (Class R) is a complex phenomenon, one aspect of which is likely to be document length.

Looking at the other classes of anaphora reveals little because, in general, few clear patterns emerge. Only two classes produced consistent positive results in both databases: the nominal substitutes (D) and the adverbs (H). No class of anaphora produced comparable negative findings. For the central pronouns (Class A), the differences are between the two databases. Engineers do not seem to use these pronouns as often or in the same manner as do writers in the social/behavioral sciences. For INSPEC, not a single query was affected, positively or negatively, by resolving these pronouns. Whereas for PsycINFO, three queries profited from the resolution of pronouns and none were adversely affected by it.

There are other differences among the databases. Appendices E and F shows that PsycINFO had twice as many positive findings as INSPEC, but both had approximately the same number of negative findings. In Table 5, we can see differences in terms of queries. Though the 12 queries from each database were selected carefully, three from INSPEC (#142, #182, #184), but only one from PsycINFO (#223) had no significant results in any anaphoric class. These differences between the databases are probably real and reflect real differences in the writing style of each field and the nature of its vocabulary.

In summary, the results indicate that a direct substitution of anaphors with their referents is not likely to improve retrieval performance of scientific abstracts. Instead, if anaphora is to be useful in retrieval effectiveness, it will have to be treated in some more complex manner than was attempted here. Some obvious treatments are discussed below.

TABLE 5

SUMMARY OF STATISTICAL RESULTS*

ANAPHORIC CLASS

INSPEC QUERIES	A	B	C	D	E	F	G	H	I	J	R
I-101										-	
I-103			+							-	
I-104										-	
I-107										+	+
I-109		-			-	-	-				
I-135						-					
I-142											
I-158				+	+	+	+	+	+		
I-170		-								+	-
I-180				+	+	+		+	+		+
I-182											
I-184											

, ,

PsycINFO Queries	A	B	C	D	E	F	G	H	I	J	R
P-203			+						+	+	+
P-207	+										
P-212	+	+	+	+			+	+		+/-	+/-
P-219			+	+							
P-221	+								-		+
P-222										-	-
P-223											
P-227								+			
P-230		-	-								
P-235		+						+			
P-248		+			-			+			
P-252			-		+	+				+	-

*Sign indicates presence of at least one finding that resolution significantly affected (positively or negatively) retrieval performance.

DISCUSSION

This study was based on what still seems to be a plausible in a document and therefore, by replacing them with their referents the WDF of important terms will be raised differentially in comparison with less important terms. Though we are pleased to find some results which support that hypothesis, we have not, as yet, been able to explain why no change was found in the majority of queries studied. Nor have we been able to determine why, for some of the queries, the results were counter to the hypothesis.

Document length is one interesting possibility for the anomalous results. Abstracts, as relatively short documents, may contain too few anaphors to effect a sizable change in WDF after resolution -- there is only an average of 4.5 anaphors in PsycINFO and 2.9 anaphors in INSPEC. Perhaps the resolution of anaphora will prove more effective on longer documents such as those found in full-text systems. Furthermore, as noted earlier, the resolution process frequently increases the length of the document -- often with non-substantive terms. These factors, combined with the sensitivity of the ranking methods to document length, may account for many of the results which ran counter to our hypothesis.

To explore the effect of document length, two further analyses can be conducted with the existing data. First, other ranking methods can be tried, ones not based on term weighting schemes or similarity measures that are sensitive to the number of tokens in the document. Second, resolutions can be automatically compared with query terms to ensure that only substantive terms are added to the resolved version of the document. Whether these analyses shed light on the various aspects of document length remains to be seen.

Another possible contributor to the unanticipated results is the form of the relevance judgment. A more continuous measure of relevance would have given more power and sensitivity to the statistical measures. When the difference between the relationships being tested in Appendix E are not statistically significant, it may be because there is no effect on resolution. However, an equal relationship can also occur when genuine differences exist. Because the users' relevance judgments were originally collected on a gross scale, measures of relationship are insensitive to differences in predicted relevance within any one of the four user-given relevance categories. Thus some of the anomalous results could be caused by a measurement limitation. Though possible, we find this explanation less plausible than that of document length.

Other explanations for the obtained results are likely to emerge from a careful study of individual retrieved documents. It is likely that some documents are strongly affected by resolution while others are not. This study examined the effect of resolution against a query and as a result "averaged out" the effect on the individual documents. A thorough analysis of what happened to individual documents within a given query should be instructive.

From Table 5 several queries seem obvious candidates for this "micro-evaluation" [18]. Query #109 is interesting because all of the significant findings were negative and there was no cumulative effect in Class R. Query #158 is similar except that the results for the individual anaphoric classes were positive. It might be instructive to compare the analysis of #158 with that of #180 (and #203) where all the positive results did lead to an overall positive effect in Class R. Query #212 is the only query that produced mixed results in the merged resolution set; perhaps something could be learned from it. Finally, it probably would be useful to examine a couple of queries that failed to achieve any significant results after resolution. Taken together, this sort of failure analysis may enable us to come to a final conclusion about the viability of our original hypothesis, or at least that version of it that pertains to abstract-length documents.

REFERENCES

1. This report is based on a paper presented at the 1986 ASIS Annual Meeting, Chicago, Illinois.

2. Karen Sparck Jones, "Index Term Weighting", Information Storage and Retrieval, 9 (1973).

3. M.A.K. Halliday, & Ruqaiya Hasan, Cohesion in English, (London: Longman Group, 1976).

4. Daniel G. Bobrow, "A Question Answering System for High School Algebra Word Problems", AFIPS Conference Proceedings, 26 (1964).

5. Terry Winograd, Understanding Natural Language, (New York: Academic Press, 1972).

6. William A. Woods, et al. The LUNAR Science Natural Language Information System: Final Report, (Cambridge, Mass: Bolt, Beranek and Newman, Inc., Report 2378, 1972).

7. Charles J. Rieger, "Conceptual Memory and Inference". In Roger C. Schank (Ed.), Conceptual Information Processing. (Amsterdam: North-Holland, 1975).

8. Yorick A. Wilks, "An Intelligent Analyzer and Understander of English", Communications of the ACM, 18 (1975) 264-274.

9. Donald E. Walker, Understanding Spoken Language, (Amsterdam: North-Holland, 1978).

10. Wendy Lehnert, The Process of Question Answering: A Computer Simulation of Cognition. (Hillsdale, N.J., Lawrence Erlbaum, 1978).

11. Gerard Salton & Michael J. McGill, Introduction to Modern Information Retrieval, (New York, McGraw-Hill, 1983).

12. John O'Connor, "Text Searching Retrieval of Answer Sentences and Other Answer Passages", Journal of the American Society for Information Science, 24 (1973) 445-460.

13. Donald Walker, "The Organization and Use of Information: Contributions of Information Science, Computational Linguistics and Artificial Intelligence". Journal of the American Society for Information Science, 32 (1981) 347-363.

14. Chris D. Paice, "The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self-Indicating Phrases". In R. N. Oddy (Ed.) Information Retrieval Research, (London: Butterworths, 1981), 172-191.

15. Elizabeth Liddy, et al. "A Study of Discourse Anaphora in Scientific Abstracts". Journal of the American Society for Information Science, in press.

16. Jeffrey Katzer, et al. A Study of the Impact of Representations in Information Retrieval Systems. (Final Report to the National Science Foundation, July 1982).

17. Michael J. McGill, et al. An Evaluation of Factors Affecting Document Ranking by Information Retrieval Systems. (Final Report to the National Science Foundation, 1979).

18. D. W. King & E. C. Bryant, The Evaluation of Information Services and Products. (Washington, D.C., Information Resources Press, 1971).

BIBLIOGRAPHY

BIBLIOGRAPHY

Belkin, N.J.; Oddy, R.N.; Brooks, H.M. "ASK for Information Retrieval: Part 1. Background and Theory. Part II, Results of a Design Study." Journal of Documentation. 38 (2,3): 61-71, 145-164; 1982.

Bobrow, Daniel G., "A Question Answering System for High School Alebra Word Problems", AFIPS Conference Proceedings, 26 (1964).

Grimes, J. The Thread of Discourse. (The Hague: Mouton Publishers; 1975.)

Halliday, M.A.K., & Ruqaiya Hasan, Cohesion in English (London: Longman Group, 1976).

Hirst, G. Anaphora in Natural Language Understanding: A Survey. New York: Springer- Verlag; 1981.

Katzer, J., et al. A Study of the Impact of Representations in Information Retrieval Systems. (Final Report to the National Science Foundation, July, 1982).

King, D.W. & Bryant, E.C. The Evaluation of Information Services and Products. Washington, D.C. Information Resources Press, 1971.

Lehnert, Wendy, The Process of Question Answering: A Computer Simulation of Cognition. (Hillsdale, N.J., Lawrence Erlbaum, 1978).

Liddy, Elizabeth, et al. "A Study of Discourse Anaphora in Scientific Abstracts". Journal of the American Society for Information Science, in press.

McGill, Michael J., et al. An Evaluation of Factors Affecting Document Ranking by Information Retrieval Systems. (Final Report to the National Science Foundation, 1979).

O'Connor, John, "Text Searching Retrieval of Answer Sentences and other Answer Passages". Journal of the American Society for Information Science, 24 (1973) 445-460.

Oddy, R.N. A Study of Representations for Anomalous States of Knowledge in Information Retrieval. NSF Grant Proposal IST-8420608. Syracuse, New York: Syracuse University, School of Information Studies: 1984.

Paice, Chris D. "The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self-Indicating Phrases". In R.N. Oddy (Ed.) Information Retrieval Research, (London: Butterworths, 1981), 172-191.

Quirk, R.; Greenbaum, S.; Leech, G.; Svartik, J. A Grammar of Contemporary English. London: Longmans; 1972.

Rieger, Charles J, "Conceptual Memory and Inference", In Roger C. Schank (Ed.) Conceptual Information Processing. (Amsterdam: North- Holland, 1975).

Salton, Gerard & McGill, Michael J. Introduction to Modern Information Retrieval. (New York, McGraw-Hill, 1983).

Sidner, C. Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. Artificial Intelligence Laboratory Report TR-537; Cambridge, Massachusetts: Massachusetts Institute of Technology; 1979.

Sparck Jones, Karen, "Index Term Weighting:, Information Storage and Retrieval, 9 (1973) 619-633.

Waldstein, R. The Role of Noun Phrases as Content Indicators. Ph.D. Dissertation, Syracuse University, School of Information Studies, 1981.

Walker, Donald E. Understanding Spoken Language. (Amsterdam: North-Holland, 1978).

Walker, Donald, "The Organization and Use of Information: Contributions of Information Science, Computational Linguistics and Artificial Intelligence". Journal of the American Society for Information Science, 32 (1981) 347-363.

Webber, B.L. A Formal Approach to Discourse Anaphora. New York: Garland Publishing, Inc.; 1979.

Wilks, Yorick A., "An Intellient Analyzer and Understander of English", Communications of the ACM, 18 (1975) 264-274.

Winograd, Terry, Understanding Natural Language. (New York: Academic Press, 1972).

Woods, William A., et al. The LUNAR Science Natural Language Information System Final Report. (Cambridge, Mass: Bolt, Beranek and Newman, Inc., Report 2378, 1972).

APPENDICES

Page

Appendix A - Preliminary Test, Functional Indexes 45

Appendix B - Retrieval Experiment Functional Indexes 55

Appendix C - Linguistic Analysis 67

Appendix D - Test Results of Rule Sets 131

Appendix E - Retrieval Test Results INSPEC Series 100 PsycABS Series 200 134

Appendix F - Summaries of Statistical Results, INSPEC and PsycABS 354

APPENDIX A

Preliminary Test,
Functional Indexes

PRELIMINARY TEST

CLASS SUMMARY

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
Central Pronouns	244	60	.80	143	46	.76	387	106	.78
Nominal Demonstrative	176	265	.40	155	148	.51	331	413	.44
Relative Pronouns	227	255	.47	192	88	.69	419	343	.55
Nominal Substitutes	60	63	.49	64	71	.47	124	134	.48
Pro-verb	21	42	.33	3	12	.20	24	54	.31
Indefinites	128	317	.29	44	209	.17	172	526	.25
Adjectives	27	51	.35	10	40	.20	37	91	.29
Adverbs	25	52	.32	30	68	.31	55	120	.31
S & Ss	188	25	.88	--	--	--	188	25	.88
Definite Article	251	1483	.14	216	1485	.13	467	2968	.14
TOTALS	1347	2613	.34	857	2167	.28	2204	4780	.32

"Central" Pronouns
(Personal, Possessive, Reflexive)

CLASS SUMMARY

PRELIMINARY TEST SET

	TERM	PSYCH. ABS.			INSPEC			TOTALS		
		Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
---PERSONAL---	he	5	-	1.00	6	-	1.00	11	-	1.00
	him	-	-	-	1	-	1.00	1	-	1.00
	I	0	1	0	0	0	0	0	1	0
	she	6	-	1.00	-	-	-	6	-	1.00
	them	12	-	1.00	3	-	1.00	15	-	1.00
	they	41	-	1.00	14	-	1.00	55	-	1.00
	us	0	1	0	0	0	0	0	1	0
	we	0	5	0	0	1	0	0	6	0
POSSESSIVE---	her	12	-	1.00	-	-	-	12	-	1.00
	his	13	-	1.00	7	-	1.00	20	-	1.00
	its	16	-	1.00	42	-	1.00	58	-	1.00
	our	0	2	0	0	2	0	0	4	0
	their	90	-	1.00	28	-	1.00	118	-	1.00
	your	0	3	0	0	0	0	0	3	0
REFLECTIVE	herself	-	-	-	-	-	-	-	-	-
	himself	2	-	1.00	-	-	-	2	-	1.00
	itself	3	-	1.00	-	-	-	3	-	1.00
	themselves	9	-	1.00	-	-	-	9	-	1.00
	Sub-total	209	12	.95	101	3	.97	310	15	.95
	it	35	48	.42	42	43	.49	77	91	.46
	TOTALS	244	60	.80	143	46	.76	387	106	.78

Nominal Demonstratives

CLASS SUMMARY

PRELIMINARY TEST SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
this	71	7	.91	88	59	.60	159	66	.71
these	61	-	1.00	50	3	.94	111	3	.97
those	29	6	.83	7	1	.88	36	7	.84
that	15	252*	.04	10	85*	.10	25	337*	.07
TOTALS	176	265	.40	155	148	.51	331	413	.44

*occurrences of "that" used as a relative pronoun are not included in this figure.

Relative Pronouns

CLASS SUMMARY
PRELIMINARY TEST SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
who	39	-	1.00	2	-	1.00	41	-	1.00
whom	3	-	1.00	-	-	-	3	-	1.00
whose	3	-	1.00	2	-	1.00	5	-	1.00
which	71	2	.97	120	2	.98	191	4	.98
where	9	1	.90	19	1	.95	28	2	.93
that	102	252*	.29	49	85*	.37	151	337*	.31
TOTALS	227	255	.49	192	88	.69	419	343	.55

*occurrences of "that" as a nominal demonstrative are not included in this figure.

Nominal Substitutes

CLASS SUMMARY
PRELIMINARY TEST SETS

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
above	-	3	-	1	-	1.00	1	3	.25
former	-	-	-	2	-	1.00	2	-	1.00
last	-	2	0	-	2	0	-	4	0
latter	3	0	1.00	4	0	1.00	7	0	1.00
one one's ones'	7	17	.29	8	30	.21	15	47	.24
other	33	8	.80	27	4	.87	60	12	.83
others	4	6	.40	6	0	1.00	10	6	.62
same	13	15	.46	4	11	.26	17	26	.40
Sub-Totals	60	51	.54	52	47	.52	112	98	.53
first	-	5	0	8	12	.40	8	17	.32
second	-	2	0	3	7	.30	3	9	.25
third	-	3	0	1	-	1.00	1	3	.25
fourth	-	1	0	-	2	0	-	3	0
fifth	-	1	0	-	1	0	-	2	0
sixth	-	-	-	-	1	0	-	1	0
seventh	-	-	-	-	-	-	-	-	-
eighth	-	-	-	-	-	-	-	-	-
ninth	-	-	-	-	1	0	-	1	0
tenth	-	-	-	-	-	-	-	-	-
Sub-Totals	-	12	0	12	24	.33	12	36	.25
TOTALS	60	63	.49	64	71	.47	124	134	.48

Proverbial

CLASS SUMMARY
PRELIMINARY TEST SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
do (do, did, does, doing, done)	21	42	.33	3	12	.20	24	54	.31

INDEFINITES

CLASS SUMMARY
PRELIMINARY TEST SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
all	14	26	.35	7	17	.29	21	43	.33
any	2	7	.22	-	15	0	2	22	.08
anyone	-	1	0	-	1	0	-	2	0
anything	-	1	0	-	-	-	-	1	0
each	29	28	.51	17	24	.41	46	52	.47
either	20	12	.62	2	3	.40	22	15	.59
enough	-	1	0	-	-	-	-	1	0
every	-	2	0	-	4	0	-	6	0
everything	-	-	-	-	-	-	-	-	-
few	-	6	0	-	4	0	-	10	0
fewer	1	9	.10	-	-	-	1	9	.10
least	-	8	0	-	6	0	-	14	0
less	10	22	.31	-	4	0	10	26	.28
little	-	4	0	-	1	0	-	5	0
many	2	10	.17	1	14	.07	3	24	.11
more	39	49	.44	11	11	.50	50	60	.46
most	2	29	.06	1	8	.11	3	37	.08
much	3	4	.43	1	2	.33	4	6	.40
neither	2	-	1.00	-	1	0	2	1	.66
no	3	58	.05	2	13	.13	5	71	.06
none	-	1	0	1	1	.50	1	2	.33
nothing	-	1	0	-	-	-	-	1	0
several	-	9	0	-	30	0	-	39	0
some	1	26	.04	1	50	.02	2	76	.02
someone	-	-	-	-	-	-	-	-	-
something	-	3	-	-	-	-	-	3	0
TOTALS	128	317	.29	44	209	.17	172	526	.25

Adjectives

CLASS SUMMARY

PRELIMINARY TEST SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
additional	-	2	0	1	3	.25	1	5	.17
another	2	3	.40	1	3	.25	3	6	.33
both	18	36	.33	1	25	.04	19	61	.25
else	-	-	-	-	1	0	0	1	0
equal	-	5	0	-	1	0	0	6	0
identical	0	2	0	0	1	0	0	3	0
TOTALS	20	48	.29	3	34	.08	23	82	.22

Pro-Adverbials

CLASS SUMMARY

PRELIMINARY TEST SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
here	-	2	0	1	2	.33	1	4	.20
identically	-	-	-	-	1	0	-	1	0
similarly	1	1	.50	-	-	-	1	1	.50
so	3	8	.27	-	15	0	3	23	.12
such	16	17	.48	20	27	.42	36	44	.45
then	1	14	.07	1	24	.04	2	38	.05
there	-	36	0	1	24	.04	1	60	.02
therin	-	-	-	1	-	1.00	1	0	1.00
thus	-	10	0	-	3	0	-	13	0
viceversa	1	-	1.00	-	-	-	1	-	1.00
TOTALS	22	88	.20	24	96	.20	46	184	.20

APPENDIX B

Retrieval Experiment,
Functional Indexes

RETRIEVAL EXPERIMENT

CLASS SUMMARY

CLASS	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
Central Pronouns	257	55	.82	188	29	.87	445	84	.84
Nominal Demonstrative	148	172	.46	169	131	.56	317	303	.51
Relatives	219	168	.57	184	82	.69	403	250	.62
Nominal Substitutes	78	76	.51	49	76	.39	127	152	.45
Pro-verb	11	20	.35	1	17	.06	12	37	.24
Indefinites	112	235	.32	35	202	.15	147	437	.25
Adjectives	27	51	.35	10	40	.20	37	91	.29
Adverbs	25	52	.32	30	68	.31	55	120	.31
S & Ss	124	25	.83	-	-	-	124	25	.83
Definite	277	1303	.17	327	1526	.18	604	2829	.18
TOTALS	1278	2157	.37	993	2171	.31	2271	4328	.34

'Central' Pronouns
(Personal, Possessive, Reflexive)

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

	TERM	PSYCH. ABS.			INSPEC			TOTALS		
		Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
Personal	he	3	-	1.00	11	-	1.00	14	-	1.00
	him	1	-	1.00	-	-	1.00	1	-	1.00
	she	5	-	1.00	1	-	1.00	6	-	1.00
	they	28	-	1.00	17	-	1.00	45	-	1.00
	them	14	-	1.00	8	-	1.00	22	-	1.00
Possessive	his	15	-	1.00	7	-	1.00	22	-	1.00
	her	13	-	1.00	-	-	1.00	13	-	1.00
	its	15	-	1.00	42	-	1.00	57	-	1.00
	their	123	-	1.00	39	-	1.00	162	-	1.00
Reflexive	herself	1	-	1.00	-	-	1.00	1	-	1.00
	himself	-	-	1.00	-	-	1.00	-	-	1.00
	itself	2	-	1.00	3	-	1.00	5	-	1.00
	themselves	9	-	1.00	3	-	1.00	12	-	1.00
	sub-total	229	-	1.00	131	-	1.00	360	-	1.00
	it	28	55	.34	57	29	.66	85	84	.61
	TOTALS	257	55	.82	188	29	.87	445	84	.84

Nominal Demonstratives

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
this	56	7	.88	88	54	.62	144	61	.70
these	52	1	.98	64	2	.97	116	3	.97
those	25	1	.96	6	6	.50	31	7	.82
that	15	163	.08	11	69	.13	26	232	.10
TOTALS	148	172	.46	169	131	.56	317	303	.51

Relative Pronouns

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
who	63	-	1.00	5	4	.55	68	4	.94
whom	3	-	1.00	-	-	-	3	-	1.00
whose	6	-	1.00	4	-	1.00	10	-	1.00
which	68	4	.94	123	1	.99	191	5	.97
where	6	62	.75	4	8	.33	10	10	.50
that	73	162	.31	48	69	.41	121	231	.34
TOTALS	219	168	.57	184	82	.69	403	250	.62

Nominal Substitutes

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
above	1	1	.50	2	-	1.00	3	1	.75
former		-	-	-	-	-	-	-	-
last	1	-	1.00	1	3	.25	2	3	.40
latter	2	-	1.00	3	-	1.00	5	-	1.00
one, ones	10	15	.40	10	25	.29	20	40	.30
other	43	9	.83	15	4	.79	58	13	.82
others	3	12	.20	2	-	1.00	5	12	.29
same	2	6	.25	6	1	.86	8	7	.53
Sub-Totals	62	43	.59	39	33	.54	101	76	.57
first, 1st	7	13	.35	5	27	.16	12	40	.23
second, 2nd	6		.75	4	13	.24	10	15	.40
third, 3rd	2	5	.29	1	2	.33	3	7	.30
fourth,4th	1	4	.20	-	-	-	1	4	.20
fifth,5th	-	2	0	-	1	0	-	3	0
sixth,6th	-	5	0	-	-	-	-	5	0
seventh,7th	-	1	0	-	-	-	-	1	0
eighth,8th	-	1	0	-	-	-	-	1	0
ninth,9th	-	-	-	-	-	-	-	-	-
tenth,10th	-	-	-	-	-	-	-	-	-
Sub-totals	16	33	.33	10	43	.19	26	76	.25
TOTALS	78	76	.51	49	76	.39	127	152	.45

Proverbials

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
Do (do, did, does,doing done)	11	20	.35	1	17	.06	12	37	.24

Indefinites

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
all	12	17	.41	2	28	.06	14	45	.24
any	1	9	.10	2	13	.13	3	22	.12
anyone	-	-	-	-	-	-	-	-	-
anything	-	-	-	-	-	-	-	-	-
each	28	14	.66	17	14	.55	45	28	.62
either	15	2	.88	6	-	1.00	21	2	.91
enough	-	-	-	-	3	0	-	3	0
every	-	2	0	-	6	0	-	8	0
everything	-	1	0	-	-	-	-	1	0
few	-	5	0	-	4	0	-	9	0
fewer	1	5	.17	-	-	-	1	5	.17
least	-	3	0	-	4	0	-	7	0
less	4	10	.28	1	2	.33	5	12	.29
little	-	9	0	-	2	0	-	11	0
many	1	9	.10	1	33	.03	2	42	.04
more	40	56	.42	3	30	.09	43	86	.33
most	3	20	.13	-	7	0	3	27	.10
much	-	3	0	2	8	.20	2	11	.15
neither	1	5	.16	-	-	-	1	5	.16
no	2	29	.06	-	4	0	2	33	.06
none	1	1	.50	-	-	-	1	1	.50
nothing	-	1	0	-	-	-	-	1	0
several	1	11	.08	1	20	.05	2	31	.06
some	2	22	.08	-	24	0	2	46	.04
someone	-	1	0	-	-	-	-	1	0
something	-	-	-	-	-	-	-	-	-
TOTALS	112	235	.32	35	202	.15	147	437	.25

Adjectives

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
additional	-	3	0	-	4	0	-	7	0
another	2	7	.22	6	2	.75	8	9	.47
both	25	32	.44	3	32	.09	28	64	.30
else	-	1	0	-	-	-	-	1	0
equal	-	2	0	-	1	0	-	3	0
identical	-	6	0	1	1	.50	1	7	.12
TOTALS	27	51	.35	10	40	.20	37	91	.29

Pro-Adverbials

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
here	-	2	0	-	2	0	-	4	-
identically	-	-	-	-	-	-	-	-	-
similarly	1	-	1.00	-	-	-	1	-	1.00
so	3	3	.50	-	19	0	3	22	.12
such	21	8	.72	27	19	.59	48	27	.64
then	-	6	0	1	9	.10	1	15	.06
there	-	29	0	-	12	0	-	41	0
therein	-	-	-	1	-	1.00	1	-	1.00
thus	-	4	0	-	7	0	-	11	0
viceversa	-	-	-	1	-	1.00	1	-	1.00
TOTALS	25	52	.32	30	68	.31	55	120	.31

Subject

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
S and Ss	124	25	.83	-	-	-	124	25	.83

Definite Article

CLASS SUMMARY

RETRIEVAL EXPERIMENT SET

TERM	PSYCH. ABS.			INSPEC			TOTALS		
	Ana.	Non.	F.I.	Ana.	Non.	F.I.	Ana.	Non.	F.I.
the	277	1303	.17	327	1526	.18	604	2829	.18

APPENDIX C

Linguistic Analysis

APPENDIX C

This appendix contains results of our detailed linguistic analysis of all the potentially anaphoric terms which were observed to function anaphorically in the sample of 600 abstracts. This analysis attempted to delineate the specific lexical environments which could be used to reliably predict whether a term was anaphoric or not in specific instances.

These rule-oriented analyses then served as the basis for the rule-sets tested by independent judges to determine whether the P.A./F.A. distinction was adequately captured in the rules. The high success rate of that testing (see Appendix D) indicates that these algorithmic type rules, once captured in formalized code, may be useful in enabling a system to determine automatically whether a P. A. is an F. A.

Contents of Appendix C

- Central Pronouns
  - 'It'
- Nominal Demonstratives
- Relative Pronouns
- Nominal Substitutes
  - 'One'
  - 'Same'
  - 'Other'
  - 'Others'
  - Ordinals
- Pro-verb 'do'
- Indefinites
  - Universals
    - 'Each'
    - 'All'
  - Multals
    - 'Many'
    - 'More'
    - 'Most'
    - 'Much'
  - Paucals
    - 'Less'
  - Assertives
    - 'Some-' group
  - Non-Assertives
    - 'Any'
    - 'Either'
  - Negatives
    - 'No'
- Residual Adjectives
- Adverbs
  - 'So'
  - 'Such'
- Subjects

Central Pronouns

Quirk & Greenbaum sub-divide the major class of 'Central Pronouns' into these three minor classes: 1) Personal pronouns; 2) Possessive pronouns, and; 3) Reflexive pronouns. The individual members of these classes are:

Personal: I, me, we, us, you, he, him, she, her,[1] it,[2] they, them.

Possessive: my, mine, our, ours, your, yours, his, her, hers, its, their, theirs.

Reflexive: myself, ourselves, yourself, yourselves, himself, herself, itself, themselves.

Only those pronouns which are underlined were observed in the subset of 600 abstracts.

The anaphoric use of these three types of pronouns can be predicted by the person distinction which pronouns demonstrate (1st person, 2nd person, 3rd person). Of the 325 occurrences of the Central Pronouns, all 1st and 2nd person pronouns (15 occurrences), whether personal, possessive, or reflexive were non-anaphoric, while all 3rd person pronouns (310 occurrences) were anaphoric.

The non-anaphoric uses are deictic references to either the author(s) of the abstract:

(1) Our research complements the EPA guidelines...

(2) The system that we are developing...

or to rather indeterminate, unspecified individuals:

(3) Discovering your radiant self. (title)

(4) Three paradoxes are considered: (a) We hurt and are hurt by those we love...

(5) Explores the idea that Gestalt concepts apply to our physical as well as our mental being.

It is unnecessary, therefore, to develop rules to determine whether in a particular instance a central pronoun is anaphoric or not. Automatic matching against lists of pronouns tagged for person-distinction should suffice to locate anaphoric references.

[1] Belongs to both Personal and Possessive classes

[2] 'It' is handled in a separate analysis.

IT

From information found in standard grammar sources,[1] it appears that 'it' has four possible uses, only two of which have been observed extensively in the samples of abstracts. The other two uses had only 1 occurrence each. All four will, however, be detailed here since it is necessary to weed out all non-anaphoric uses. The first three uses are non-anaphoric and they will be presented in order of their ease of distinguishability from the other cases. The anaphoric use will be presented last since it does not occur in as predictable a syntactic environment as do the nonanaphoric uses.

Empty/Prop: 'It' may be used to refer to the rather indeterminate notion of the general state of affairs. Frequently this use is to do with the weather or the time.

(2) It is raining out.

(3) It is nine-thirty.

The next two uses appear to be special cases of the more general notion of cataphoric use of 'it'. In both uses the referent for which 'it' is substituting, follows 'it' in the text.

Anticipatory: 'It' appears as the result of rearrangement of terms from the usual S-V-O word order by the movement rule known in transformational grammar as extraposition. This involves movement of a clausal subject from the original syntactic structure of:

clausal subject + pred

to a position toward the end of the sentence. The postponed element's position is filled by the anticipatory pronoun 'it'. The resulting syntactic structure is:

'it' + pred + clausal subject

Observation of the abstracts reveals that the predicate in this type construction appears to be either:

1. Of the class of cognitive/emotive verbs of thinking, knowing, feeling, etc., followed by 'that' and an independent clause:

(4a) It is emphasized that the evidence was obtained from normal children reared in their natural homes by their biological parents.

---

[1] Quirk, Greenbaum, Leech & Svartvik. A Grammar of Contemporary English. Longman Group, 1980.

2. A Form of 'to be' followed by an adjective plus 'to' or 'for'. Some common constructions are: 'It it possible for': 'It was difficult for': 'It is unrealistic to'.

> (5a) It is crucial for therapists to feel free to discuss uses and abuses of this money with patients.

In all instances of extraposition, one can easily rearrange the sentence elements to return to normal S-V-O order by substituting the clausal subject for 'it'.

> (4b) That the evidence was obtained from normal children reared in their natural homes by their biological parents is emphasized.

> (5b) For therapists to feel free to discuss uses and abuses of this money with patients is crucial.

Cleft sentence: 'It' is used in constructions of this type to permit focal prominence to be given to a particular item in the sentence. Sentence elements are rearranged from normal order to:

'It' + form of 'to be' + focus element + rel. clause

> (6) It was the weather that caused the picnic's cancellation.

Cleft sentences can be differentiated from anticipatory constructions by the fact that the clause postponed in anticipatory usage is an independent clause in which the subordinating conjunction does not fill a syntactic slot. On the other hand, in cleft sentences the head of the relative clause fills a syntactic role in the clause.

Anaphoric: 'It' performs as an anaphoric item when 'it' is in its role as a personal pronoun, i. e., it serves as an abbreviated reference to a more fully explicated antecedent. However, 'it' differs from all other personal pronouns in that 'it' has the capability of extended reference. 'It' may replace a whole clause or sentence or 'it' may simply refer to a single word. Also, in anaphoric usage, 'it' may be related to its antecedent either by 'identity of reference' or 'identity of specification'. In identity of reference, 'it' refers to the exact same entity as the antecedent.

> (7) Feedback has an impact on the strength of beliefs to which it is targeted.

Whereas, in identity of specification, 'it' refers to a separate entity but one that is specified in same manner as its antecedent.

(8) Paul ordered a dish of spaghetti for dinner.
Bill ordered it too.

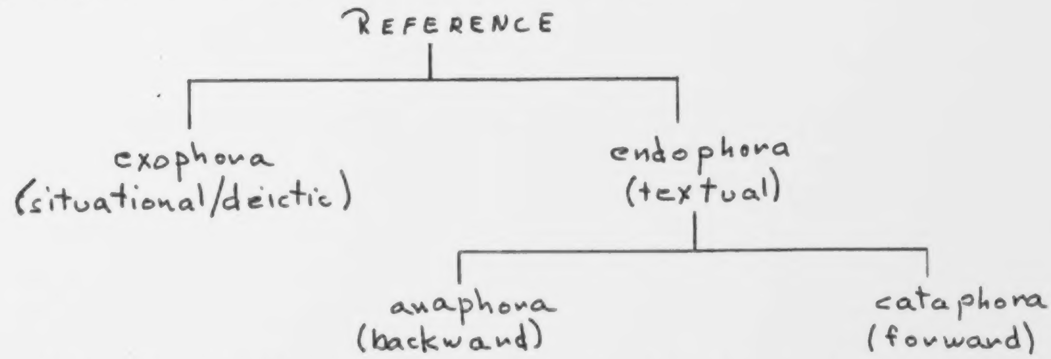
Nominal Demonstratives

Demonstrative reference is essentially a form of verbal pointing. There are 4 nominal demonstratives: 'this', 'these', 'those', and 'that'. The nominal demonstratives 'this', 'these', and 'those' function only as referential items. They have no other use. 'That' has four senses associated with it, and will be treated separately.

THIS, THESE, THOSE

When any of these 3 terms are encountered in the text, what must be determined, therefore, is:

1. Whether the reference is situational (exophoric/deictic) or textual (endophoric).
2. If endophoric, whether the reference is backward in text (anaphoric) or forward in text (cataphoric).



In making the first determination, the fact that abstracts are quite self-contained and non-situationally dependent predicts that the endophoric use is common and the exophoric quite uncommon. This has been observed to be the case. Therefore, it is more efficient to proceed by determining the contextual clues (lexical, not syntactic) that indicate exophoric use, rather than clues to endophoric use.

There exist two general cases of exophoric use of nominal demonstratives as exhibited in abstracts. The first of these is deictic reference to either: 1) the document of which the abstract is a part or: 2) the time at which the document was written. 'This' is the usual nominal demonstrative chosen for such use and typical phrases are: 'in this paper', 'at this writing', or 'this report'.

> E.g. "The discussion of sexual behavior in this paper is confined to heterosexual activities."

The second general case of exophoric use of nominal demonstratives exemplifies the larger phenomenon of refering to indeterminate referents which are presumed to exist but which are not specified. Phrases composed of double pronouns such as 'those who', 'anyone who', or 'that which' are common lexical indicators used to refer to someone orsomething without actually denoting anyone oranything.

> E.g. "People who buy social science should remember that data can easily be misconstrued or misrepresented by those who wish to prove their particular argument, for any of a number of reasons."

The second determination is whether the endophoric reference is anaphoric or cataphoric. In abstracts, cataphoric noun phrases are used to introduce a list, and are usually followed by a colon.[1]

> E.g. "The experiment tested these three approaches:"

Classifying

Having eliminated the non-functioning P.A.'s, the F.A.'s may be classified. 'This', 'these' and 'those' function anaphoricaly either as:

- Demonstrative adjective
- Demonstrative pronoun

Classify as demonstrative adjective if the term is followed by a noun or an adjective. Otherwise, classify as demonstrative pronoun.

THAT

'That' has four senses associated with it. Three of these are referential uses and the reference is anaphoric for each use.

- Demonstrative adjective
- Demonstrative pronoun
- Relative pronoun

Non-anaphorically, 'that' functions as:

- Subordinating conjunction

To determine in a particular instance whether 'that' is an F.A., the one non-anaphoric use will be tested for, and all such uses excluded from further analysis.

'That', in its role as a subordinating conjunction, occurs in two contexts and appears to be acting as a lexical colon in both. In one context, the role of 'that' as a subordinating conjunction is recognizable by these two facts.

1. 'That' follows cognitive/emotive verbs of:

   - knowing
   - thinking
   - believing
   - fearing
   - saying
   - remembering
   - perceiving

   or their nominalisation:

   - assumption
   - suggestion
   - hypothesis
   - explanation
   - suggestion

2. The clause introduced by 'that' contains no empty syntactic slot, i. e. the clause is complete, it consists of subject-verb-object, in any order.

   > E.g. "It was determined that a very definite advantage is achieved when the airflow is reversed periodically."

One seemingly troublesome construction, 'that is', is actually an ellipsed variant of the phrase 'that is to say' and serves as an indicator of a subsequent phrase of apposition. The ellipsed verb 'say' belongs to the class of cognitive verbs which indicate use of 'that' as a subordinating conjunction. Therefore, the construction 'that is' will be classified as such.

> E.g. "In the first, the concern is to construct a resistivity structure whose responses are acceptably close to the observations, that is, the measured amplitudes and/or phrases."

In the second context, 'that' is one component of a compound subordinating conjunction, and is recongnizable by two facts:

1. 'That' is the final element in constructions of the following type:

   - but that
   - in that
   - such that
   - so that
   - in order that

2. Again in this context, the clause introduced by 'that' contains no empty syntactic slot.

   > E.g. "Skinner's concept of contingencies of reinforcement may be a crucial one for understanding the relationship between the arts and the sciences in that each involves processes and products of human behavior."

Classifying

Having excluded non-anaphoric occurrences of 'that', the remaining instances may be classified.

- Classify as relative pronoun if 'that' introduces a clause that is not complete, i.e. contains a syntactically empty slot.

  > E.g. "Phillips developed a system that diagnosed human illness."

- Classify as demonstrative adjective if followed by noun or adjective.

  > E.g. "Selected components of that framework are empirically tested."

- Otherwise, classify as demonstrative pronoun.

  > E.g. "The performance of the model is compared to that of the physicians."

[1] Based on three cataphoric instances in sample of 600 abstracts.

Relative Pronouns

Relative pronouns introduce relative clauses postmodifying nominal heads, and have anaphoric reference to the antecedent noun phrase which is postmodified by the entire relative clause. This class of pronouns consists of the following terms: who, whom, whose, which, and that.[1] All of these terms were observed in the subset of 600 abstracts.

All occurrences of 'who', 'whom', and 'whose' were anaphoric while 4 of the 195 ocurrences of 'which' were nonanaphoric. Therefore, only rules for determining anaphoric vs. nonanaphoric use of 'which' were developed.

Anaphoric Use

The anaphoric use of 'which' occurs in three different syntactic environments.

1. 'Which' may follow immediately the nominal head it postmodifies.

   (1) Performance is compared with the traditional algorithm which employs only swapping.

2. 'Which' may be immediately preceded by and function as object of a preposition.

   (2) The process is modelled by a hyperbolic system in which the inflows act both as distributed and as boundary controls.

3. 'Which' may follow a verb in the passive voice, which separates the relative pronoun from the nominal head it postmodifies. This usage can be determined by the fact that these passive verb phrases can be moved to the end of the relative clause without altering the meaning of the sentence or damaging its grammaticality.

   (3) An algorithm is presented which maps patterns from a high-dimensional space to a plane.

Non-anaphoric Use

- 'Which' in its nonanaphoric usage acts as an indefinite determiner of the noun phrase which follows it. The typical syntactic environment for this usage is:

  verb + 'which' + noun phrase

The verb phrase is usually active and can in no way be moved without damaging the grammaticality and sense of the sentence.

(4) The study will attempt to determine which method of analysis will be most cost-effective.

[1] That was treated in the analysis of nominal demonstratives, and will not be reconsidered here.

Nominal Substitutes

The set of terms considered as nominal substitutes was completed with the following summary results:

	P.A.		Ins.		Total		
	Ana.	Non.	Ana.	Non.	Ana.	Non.	F.I.
above	0	3	1	0	1	3	.25
former	0	0	2	0	2	0	1.00
last	0	2	0	2	0	4	0
latter	3	0	4	0	7	0	1.00
one	7	17	8	30	15	47	.24
other	33	8	27	4	60	12	.83
others	4	6	6	0	10	6	.62
same	13	15	4	11	17	26	.40
Total	60	51	52	47	112	98	.53

Obviously, the terms 'former' and 'latter' which have an F.I. of 1.00 will not be tested; nor will 'last' which had an F. I. of 0; nor will 'above' which had only 1 anaphoric use in the set of 600 abstracts.

The remaining four terms - 'one', 'other', 'others' and 'same' have separate rule sets for each term.

ONE

The term 'one' (including the forms 'ones' and 'one's') has three major senses associated with it. 'One' may be used as: 1) a numeral; 2) a nominal substitute, or; 3) an indefinite pronoun. To determine which of these senses is intended in a piece of text, it is first necessary to understand the detailed structure of a nominal group. (a.k.a. noun phrase)

NOMINAL GROUP

logical structure	Premodifiers			Head	Post-modifier
word classes	determiner	numeral	adjective	noun	prepositional phrase
	a	b	c	d	e

(1) the six red onions on the table
a b c d e

(2) the difficult ones
a c d

(3) one method
b d

(4) one current technique
b c d

(5) that smoking gives one cancer
d

The slot in which the term 'one' occurs within the nominal group will determine which use of 'one' is intended.

Nominal substitute

If 'one' functions as head (d) of a nominal group premodified by either a determiner (a), e. g. definite article or nominal demonstrative, or an adjective (c) or both, as in (2), the term is being used anaphorically as a nominal substitute. The syntactic environment would be:

premodifier(s) + 'one'

(6) The store had no gold bracelets; just silver ones.
c d

Indefinite pronoun

In its use as an indefinite pronoun, 'one' is non-anaphoric in that there is no presupposition of a more specified antecedent to which 'one' is referring. Its meaning is that of an indeterminate, generic person who cannot be defined any more specifically within the text.

(7) One never knows what might happen.
d

In terms of the nominal group structure, the indefinite pronoun 'one' has been observed in this data set to occur as the unmodified head (d), as in (5). The form 'one's' is found only in this usage.

Numeral

The most frequent use of 'one' is as a cardinal number. In some instances this sense of 'one' is non-anaphoric. In others, when its use is combined with the linguistic technique of ellipsis, it is anaphoric. The easiest non-anaphoric structure to recognize is the hyphenated combination.

> (8) One-sided sequential tests for the mean of an exponential distribution are proposed.

The remaining occurrences of 'one' is its use as a numeral can be detected by again referring to the nominal group structure. In its un-hypenated numeral uses, 'one' functions as a premodifier in a nominal group, as in (3) and (4). The structural environment would be:

'one' + (adjective) + (head) + (prepositional phrase)

In other words, when used as a numeral, 'one' is not preceded by another premodifier, but must be succeeded by at least one and possibly even all of the following:

adjective - c
head noun - d
prepositional phrase - e

(9) A control function is proposed for one possible
system configuration.
c d b c

(10) The conjecture is shown to be true for one level
of 'next' statement.
e b d

(11) The evaluation of textbooks using one of the
standard readability formulas is a lengthy task.
e b

The only exception to this rule as observed in the 600 abstracts, w-as the restrictive adjective 'only' which preceded 'one' twice in the data set, although 'one' was being used as a numeral.

To then determine whether this usage of 'one' is anaphoric or not, the prior text must be scanned for an earlier occurrence of the head noun which 'one' is modifying. If that head noun is specified in greater detail in a prior usage, then 'one' is to be considered anaphoric, since its usage establishes an acceptable environment for some premodifiers to be ellipsed.

(12a) This is illustrated by a detailed examination
of two simple microprocessor-based gaging systems.
One system measures location.
b d

When the anaphoric 'one' is resolved in this usage, the ellipsed premodifiers are re-inserted.

> (12b) This is illustrated by a detailed examination of two simple microprocessor-based gaging systems. $\underset{b}{\text{One}}$ $\underset{c}{\text{simple}}$ $\underset{c}{\text{microprocessor-based}}$ $\underset{c}{\text{gaging}}$ $\underset{d}{\text{system}}$ measures location.

'One' has been observed in this data set to be anaphoric only in the environment:

'one' + head noun

although the inverse of this is not true. That is, all instances of 'one' in this environment are not anaphoric.

'SAME'

'Same' occurred in the 600 abstracts a total of 43 times with 17 of these occurrences being anaphoric for an F. I. of .40. 3 of the 4 syntactic environments in which 'same' was observed are always non-anaphoric, while the status of 'same' in the 4th environment depends on prior text.

Non-Anaphoric

1.

'the' + 'same' + preposition

'or'

'the' + 'same' + .

1. Her responses remained the same throughout the interrogation.

2.

'same' + noun

2. The students interviewed were a very homogenous group - same likes, same dislikes.

3.

'the' + 'same' + adjective + noun

3. The majority of respondents indicated an interest in the same leisure-time activities.

Dependent on Text

When the following syntax is encountered:

'the' + 'same' + noun

'same' is non-anaphoric if the noun it pre-modifies, or that noun's synonym, was not used earlier in text in a more fully amplified reference.

On the other hand, 'same' is being used anaphorically if the noun it pre-modifies was specified earlier in text in fuller detail. The earlier specification may be in in the form of 1 or more pre-modifiers of the noun, which are ellipsed when 'same' is used in the current reference.

4. Expert searchers used the full-text approach. Novice searchers used the same approach.

Or, the noun used with 'same' may be a rather general term which was explicated earlier in more detail by either a prepositional phrase:

5. Freshmen were most concerned with the problem of having to choose a major. Some sophomores remained disturbed by the same problem.

Or, the earlier reference may have been a detailed explanation not even containing the same general term or its synonym.

6. 15 Subjects were exposed to the stimulus for 4 minutes while 15 Subjects were exposed to the control condition for the same interval.

'Other'

'Other' occurred in the 600 abstracts a total of 72 times with 60 of these occurrences being anaphoric for an F. I. of .83.

1. The basic use of 'other' is to make some kind of a comparison, but in most instances the comparison is not as fully spelled out as the underlying meaning intends. The most typical comparisons are of the form:

   (1a) This beer is sold in the U. S. and 14 other countries.

which would be resolved by moving 'other' to a position following the noun, adding the explicit comparative term 'than', and that which is being compared:

(1b) This beer is sold in the U. S. and 14 countries other than the U. S..

The typical syntax for this use would be:

'other' + (adjective) + noun

and the use is anaphoric in almost all instances except those few where there is no information given as to what the 'other' entity is being compared to:

(2) This beer is sold in 14 other countries.

There are 3 additional possible syntactic environments for the anaphoric use of 'other'.

2. 'Other' may combine with 'each' in a reciprocal reference:

   'each' + 'other'

3. 'Other' may be used as a pronominal in the following syntax:

   'the' + 'other' (not followed by a noun)

   (3a) There are two transformational grammar approaches. The first builds on Chomsky's work and the other follows Postal's model.

which would be resolved as:

(3b) There are two transformational grammar approaches. The first transformational grammar approach builds on Chomsky's work and the other transformational grammar approach follows Postal's model.

4. When the explicit comparatives 'than' or 'while' precede the noun phrase containing 'other', again a comparison is being made, but one which would be resolved differently than is the case in #1.

'than'/'while' + 'the' + 'other' + noun phrase

(4a) Groups of cats, dogs, and rabbits were exposed to the same stimulus. Dogs performed better than the other groups.

which would be resolved as:

(4b) Groups of cats, dogs, and rabbits were exposed to the same stimulus. Dogs performed better than cats and rabbits.

As was pointed out in #1, there is one syntax in which 'other' may be either anaphoric or non-anaphoric, but there is only 1 soley non-anaphoric syntax for 'other':

'other' + 'than'

(5) Universities other than S. U. have an over-emphasis on sports.

'OTHERS'

'Others' appeared in the 600 abstracts a total of 16 times with 10 of these uses being anaphoric for an F. I. of .62.

Non-Anaphoric

When used non-anaphorically, 'others' refers to indefinite individuals whose specific identity is of no concern. The non-anaphoric use of 'others' almost always follows prepositions:

(1) Concern for others is not highly valued in this society.

Anaphoric

When used anaphorically, 'others' serves as a pronominal substitute for individuals or items referred to earlier; perhaps even enumerated and has the meaning of 'more like the above'.

(2) Ss exhibited the defense mechanisms of denial, projection and others.

In this anaphoric use, 'others' either follows 'and' or functions as subject or direct object of the sentence.

ORDINALS

The ordinals, which are grouped with the nominal substitutes in this study, from 'first' to 'tenth' were observed in the 600 abstracts as follows with an overall F. I. of .25.

	Psych Abs		Inspec	
	Ana.	Non.	Ana.	Non.
first	-	5	8	12
second	-	2	3	7
third	-	3	1	-
fourth	-	1	-	2
fifth	-	1	-	1
sixth	-	-	-	1
seventh	-	-	-	-
eighth	-	-	-	-
ninth	-	-	-	1
tenth	-	-	-	-
Totals	-	12	12	24

Non-Anaphoric Use

1. Hyphenated terms in which one term is an ordinal are always nonanaphoric. Some common uses of this type are: 'second-graders', 'first-order calculus', 'one-sixth'. It does happen infrequently that the hyphen is omitted, but the notion intended by the two terms is obviously that of a known hyphenated term.

2. Titles of meetings, books, etc. frequently use ordinals nonanaphorically, e. g. 'Second Edition', 'Eighth Annual Meeting'. Ordinals are also used in less formal titles such as 'fifth generation computers'.

3. The ordinal 'first' functions nonanaphorically as an adverb with the meaning of "before another in time or space or action". Typical syntax for such a use is:

   auxiliary verb + 'first' + main verb

   (1) Subjects were first tagged and then released to the environment.

   or

   'at first'

   (2) At first, both techniques appeared to work.

   or

   'First' + complete clause

(3) First, wash your hands.

Anaphoric Use

Ordinals are always anaphoric when they are intended as the numerative adjective modifying a noun but the noun has been ellipsed and the ordinal therefore functions as the head of the noun phrase. Syntax for such a use would be:

'the' + ordinal (not followed by a noun or adjective)

(4a) Two consumer-oriented evaluation techniques were tested. The first was tried out on suburban housewives.

which would be resolved as:

(4b) Two consumer-oriented evaluation techniques were tested. The first consumer-oriented evaluation technique was tried out on suburban housewives.

Use Dependent on Text

Ordinals used as numerals in a noun phrase may or may not be anaphoric depending on whether the noun in the phrase has been expressed any more fully in prior text. In the uses observed in the 600 abstracts, all instances of the following syntax where there is an adjective between the ordinal and the noun were non-anaphoric uses.

determiner + ordinal + adjective + noun

(5) The second busiest airport is J.F.K. Airport in New York City.

Those instances in which the ordinal directly precedes the noun it modifies tend to be anaphoric but there are a few exceptions. So when the syntax:

determiner + ordinal + noun

is encountered, prior text will have to be evaluated to see whether the use is anaphoric or not.

(6a) There had been three attempts at in vitro fertilization. The third attempt was successful.

which would be an anaphoric use of an ordinal and would be resolved as:

(6b) There had been three attempts at in vitro fertilization. The third attempt at in vitro fertilization was successful.

'DO'

The only true pro-verb in the English language is the verb 'do'. In the 600 abstracts analysed, the verb 'do' appears in all 5 of its possible forms: 'did', 'do', 'does', 'doing' and 'done'. The rules for recognition of anaphoric vs. non-anaphoric use of the verb are written to encompass all forms. When the term 'do' is used in a rule it is to be interpreted as implying all of the possible forms of 'do'. On the other hand, the negative contractions of 'do' will be handled later in the verbal ellipsis class of anaphora, in that the only anaphoric use of these contractions is the elliptical one.

Non-anaphoric

The verb 'do' has two distinct non-anaphoric uses:

1. Lexical verb - meaning 'to perform' or 'to carry out'. It is always transitive (takes a direct object).

   (1)The subjects did three sets of problems.

   When the past participle form of the verb (done) occurs in this usage, the sentence is in the passive voice and the direct object will precede the verb.

   (2) The assignment was done separately by each of the students.

   When the form 'do' occurs in this usage, the sentence is frequently imperative.

   (3) Do your homework!

2. Periphrastic auxiliary - in this usage, 'do' has no individual meaning but serves as a necessary verbal operator, a purely grammatical element which is required for forming certain cases of a verb, or is added as emphasis in other instances. Periphrastic means to be formed by the use of auxiliaries instead of by inflection of the verb. Compare,

   (4a) She left.

   (4b) She did leave.

   'Do' as a periphrastic auxiliary is used when the main verb is in the simple present or past tense in the following contexts:

   - Interrogative

     (5) Did he stay long?

   - Negative

(6) Dietary treatment did not effect total volume intake.

- Marked/emphatic positive

(7) He did ask her for some assistance.

Anaphoric

The verb 'do' has 3 types of anaphoric usage:

1. Predicate substitute - the verb 'do' can be used to replace a verb or verb clause. In the genre of abstracts this use has been observed to occur in the second of two semantically contrastive clauses conjoined by a comparative term such as 'than' or 'as', and to be followed immediately by the noun phrase which is actually subject of the verb for which 'do' is substituting.

   (8a) Freshman reported less change than did seniors.

   which would be resolved as:

   (8b) Freshmen reported less change than seniors reported change.

2. Ellipsis - verbal ellipsis is actually a special case of predicate substitution where zero substitution occurs rather than lexical substitution. Use of 'do' in verbal ellipsis is decipherable in those sentences where 'do' is retained in its role of periphrastic auxiliary but the main verb is ellided.

   (9a) I don't like cheese now but I did when I was a child.

   (9b) I don't like cheese now but I did like cheese when I was a child.

   The structural environment differs from that of predicate substitution in that 'do' is not followed by the noun phrase which serves as the subject of that verb clause.

3. Complex pro-verb - when combined with 'it', 'so', 'the same', 'this' or 'that', the resulting phrases ('do it', 'do so', 'do the same' , 'do that', 'so doing' and 'do this') function as compound referential verbal groups which together replace an entire predication.

   (10) Paul woke up early, had a good breakfast, and left on time for work. Michael did the same.

INDEFINITES

Of the 33 terms considered by Quirk & Greenbaum to be indefinite pronouns, 25 were observed in the set of 600 abstracts. Of these 25 terms, 14 functioned anaphorically at least once. Therefore, rules to determine whether a term is functioning anaphorically in a specific instance were written for only these 14 terms. The table below provides summary statistics of the indefinite pronouns.

TERMS	PsychAbs.		INSPEC		TOTALS		
	Ana.	Non.	Ana.	Non.	Ana.	Non.	F.I.
UNIVERSALS							
each	29	28	17	24	46	52	.47
all	14	26	7	17	21	43	.33
every	-	2	-	4	-	6	0
everything	-	-	-	-	-	-	-
UNIVERSALS TOTALS	43	56	24	45	67	101	.40
ASSERTIVES							
many	2	10	1	14	3	24	.11
more	39	49	11	11	50	60	.46
most	2	29	1	8	3	37	.08
much	3	4	1	2	4	6	.40
few	-	6	-	4	-	10	0
fewer	1	9	-	-	1	9	.10
little	-	4	-	1	-	5	0
least	-	8	-	6	-	14	0
less	10	22	-	4	10	26	.28
several	-	9	-	30	-	39	0
enough	-	1	-	-	-	1	0
some	1	26	1	50	2	76	.02
someone	-	-	-	-	-	-	-
something	-	3	-	-	-	3	0
ASSERTIVES TOTALS	58	180	15	130	93	310	.19
NON-ASSERTIVES							
any	2	7	-	15	2	22	.08
anyone	-	1	-	1	-	2	0
anything	-	1	-	-	-	1	0
either	20	12	2	3	22	15	.59
NON-ASSERTIVES TOTALS	22	21	2	19	24	40	.38
NEGATIVES							
no	3	58	2	13	5	71	.06
none	-	1	1	1	1	2	.33
nothing	-	1	-	-	-	1	0
neither	2	-	-	1	2	1	.66
NEGATIVES TOTALS	5	60	3	15	8	75	.10
GRAND TOTALS	128	317	44	209	172	526	.25

'EACH'

'Each' is the second indefinite pronoun of the universal subclass to be considered. 'Each' is similar to 'all' in that its anaphoric use can be determined by syntax only some of the time. In the remaining instances, it is the prior text that will determine whether its use is anaphoric or not.

'Each' has three anaphoric uses:

- 'each other' functions as an anaphoric reciprocal pronoun.

  (1) The sensitized Ss were more likely to initiate conversation with each other than with nonsensitized Ss.

- 'each' functions as the head of a nominal group and in this use has been observed only in the following syntactic environment:

  'each' + verb form

  (2) 117 first-grade children were tested on the apparatus and the first two trials completed by each were recorded.

- 'each' + preposition other than 'of' (e.g. at, under, within)

  (3) Ss were 24 children, 12 each at the two levels tested.

The one syntactic environment in which 'each' invariably functions non-anaphorically is:

'each of'

Although the noun phrase following 'each of' is itself frequently anaphoric (e.g. 'each of these', 'each of which'), the term 'each' serves as a nonanaphoric quantifier meaning 'each and every one of the following entities'.

(4) Each of these functions is described in detail.

In its remaining occurrences, 'each' functions as a determiner in either of two syntactic environments:

'each' + noun

'each' + adjective + noun

In these environments, the prior text must be consulted to see whether the noun has been more fully specified in an earlier occurrence. 58 of the 98 occurrences of 'each' in the 600 abstracts are of this type which requires more than recognition of a particular syntax. The situation is further complicated by the fact that the noun which 'each' serves as a determiner for, may not be the same word as used originally, but rather a paraphrase, a semantically related word such as a synonym or general noun.

> (5) Surveyed 1,689 adult married females to examine media-exposure patterns. Each respondent was classified as....

If the noun that 'each' is serving as a determiner for, is a paraphrase of, or repetition of a more fully specified noun, then 'each' is serving an anaphoric function. Otherwise, not.

An exception to this rule is the noun phrase 'each S' since 'S' will be judged anaphoric/nonanaphoric in its own right, and each therefore serves simply as a nonanaphoric quantifier.

'ALL'

'All' is an indefinite pronoun of the subclass termed universal. Its basic definition is "every member or individual component of". 'All' has 3 basic uses:

1. When occurring in the phrase 'all that', the reference is non-anaphoric in that it is an indeterminate reference to entities which are presumed to exist but are not specified, much the same as other double pronouns such as 'those who' or 'that which' have indeterminate reference.

   (1) All that was needed was provided by the instructor.

2. 'All' functions anaphorically as an independent nominal head, in the following syntactic environments:

   - 'all' + verb form

     (2) 13 retarded children and 14 children with average IQ's were tested. All were administered the same pretest.

   - 'all' + prepositions other than 'of' (e.g. 'but', 'under', 'within')

     (3) The algorithms developed are all within the capabilities of the current system.

   - 'all' + adjective not followed by a noun

     (4) Paradoxically, suggestions for eliminating the delivery service, improving the service, or updating its mode were helpful to consider and all reasonable from the financial point of view.

3. 'All' may function as an element other than head of a nominal group. As such, 'all' may be either anaphoric or non-anaphoric based in some instances on which elements of the nominal group follow the term, and in other instances on the prior text.

   Firstly, 'all' functions non-anaphorically when it occurs as a predeterminer/quantifier in nominal groups of the following structures:

   - 'all' + 'of' + noun phrase

     (5) All of the test results were distributed first to the program coordinator.

   - 'all' + determiner (e.g. 'the', 'this', 'such', 'their') + noun phrase

     (6) All their work was for nought.

- 'all' + adjective + noun

  (7) All necessary adjustments were worked out prior to the test run.

In the following 2 nominal group structures, 'all' may be either anaphoric or nonanaphoric and the decision as to which, will be based on the semantics of the preceding text:

- 'all' + noun

  (8) All books were returned to the library prior to the new semester.

- 'all' + numeral + noun phrase

  (9) All 50 states have their own welfare assistance programs.

Where ['all' + noun] is the structure, if this is the first occurrence of the noun, 'all' will be a nonanaphoric quantifier and likely a rather generic reference, such as "all men". If it is not the first occurrence of the noun and the noun is more specified (either by premodifiers or postmodifiers) in a prior occurrence in the text, then the use is anaphoric. However, if the noun is not any more fully specified in prior use(s), then it is nonanaphoric.

In that 'all' is what is known as a congretory quantifier, it appears to perform as an anaphoric direction to readers to reassemble and enumerate all subgroups that may have been separated out in prior text. This occurs most frequently when 'all' precedes 'Ss' or general nouns such as "all groups" or "all 4 categories".

(10) 32 Ss were assigned to either progressive relaxation (PR), clinically standardized meditation (SM), or a waiting list control group (CG). At the end of a 5 week period all Ss were exposed to 6 very loud tones. All 3 groups exhibited higher heart rates.

In the above example, both "all Ss" and "all 3 groups" would be resolved by reiterating the 3 groups into which the Ss had been subdivided. The prior text will dictate whether the 'all' is anaphoric or not, for in some instances the Ss will not have been subdivided and therefore only the term Ss is anaphoric, e. g.:

(11) Investigated the possible influence of 48 hours of sleep deprivation (SD) in 12 19-30 year old males. Following SD, all Ss showed marked reduction of DNA synthesis.

'MANY'

'Many' is an indefinite pronoun of the multal subclass. It occurred in the 600 abstracts a total of 27 times with only 3 of these instances anaphoric for an F.I. of .11.

Although there exist several possible uses of 'many', only rules for the one observed use will be included here since our rule-writing is data-driven rather than theory-driven.

The only observed use of 'many' was as an adjective with the meaning - "a large but indefinite number". In this use, 'many' was observed in three different syntactic environments. In the first two, the observed uses were always nonanaphoric:

'many' + adjective + noun

(1) Decisions were made based on many previous cases.

and

'many' + 'of'

(2) Males have many of the same characteristics as females.

In the third observed environment:

'many' + noun

the prior text must be checked to see whether the noun that 'many' is modifying is specified previously in any greater detail.

(3a) Research was conducted on a variety of response-specific stimuli. Many stimuli were found to be more effective on immature cells than on fully developed ones.

(3b) Research was conducted on a variety of response-specific stimuli. Many response-specific stimuli were found to be more effective on immature cells than on developed ones.

'MORE'

With analysis of the term 'more', we encounter for the first time consideration of those types of words which serve as clues to ellipsis rather than serve as anaphors themselves. All of the terms we are analyzing in this project, when they function in a way of interest to us, will function either as:

1. Terms which are lexical anaphors, that is, place-holders to be replaced by terms used in prior text. Pronouns and nominal substitutes are prime examples.

2. Terms which serve as clues to the fact that words have been ellipsed in text. The term which serves as the clue is not itself replaced, but portions of the prior text are added to the sentence containing the clue word.

'More' is an indefinite pronoun of the multal subclass, which it shares with 'many', 'most' and 'much'. 'More' was observed in the 600 abstracts a total of 109 times. 49 of these occurrences were anaphoric for an F. I. of .45.

In all its uses, presence of the term 'more' implies the basic notion that a comparison of some type is being made. The type comparison being made will determine whether the use is always anaphoric; always nonanaphoric; or dependent on the specifics of the text.

DEPENDENT ON TEXT

Clausal

The most common comparison is between two clauses. The cooccurence of 'more' and 'than' within the same sentence establishes the necessary environment for clausal comparison although 'more' and 'than' need not be contiguous. 'More', which is considered the comparative element, together with 'than' forms a hinge by which the two clauses coalesce to form a comparative construction. The two clauses are intended to be semantic equivalents with the exception of one element which provides the contrast or comparison between the two clauses. The two clauses are closely parallel, both in structure and content. As a result, it is common practice to elide rather than repeat some portion of what the second clause has in common with the first clause. If there is this ellipsis, then for our analysis, 'more' is to be attributed with being the lexical trigger for the ellipsis. The term 'more' itself is not replaced with a term, but it serves as a structural clue that a clausal comparison is being made and that the structure of both clauses should be parallel.

Therefore, when 'more' and 'than' co-occur in a sentence, that sentence's two clauses must be compared to check whether the structures of the two clauses are completely parallel or whether some terms have been ellipsed. If there has been some ellipsis, which syntactic items have been elided may vary. For example, in sentence (1a) the verb and object of the second clause have been ellipsed:

(1a) It was found that firstborns showed more death threat than lateborns.

(1b) It was found that firstborns showed more death threat than lateborns showed death threat.

while in (2a) the subject and verb have been ellipsed:

(2a) Those with depression were more likely to have received diazapam than antidepressants.

(2b) Those with depression were more likely to have received diazapam than those with depression were likely to have received antidepressants.

Note that in resolving the ellipsis the term 'more' is not carried forward and re-used with the other terms in the 2nd clause.

However, it is not to be assumed that all sentences with co-occurrences of 'more' and 'than' have some elements elided, but rather the presence of those terms requires that the structure be checked for exact parallel construction.

It does occur somewhat infrequently (4 out of 46 ellipses) that the ellipsis appears to be both cataphoric and anaphoric, with some words from prior text and some words from later text used to flesh out a completely parallel structure. We will consider these occurences anaphoric in that both the anaphoric and cataphoric ellipses need be resolved.

(3a) The examples given indicate that younger Ss made more false than true conclusions.

(3b) The examples given indicate that younger Ss made more false conclusions than younger Ss made true conclusions.

When comparing the clauses for parallel structure, all other anaphors must be resolved first to insure that two different words are not credited with creating the same elliptical situation. This is particularly important if the verb of the second comparative clause is a form of the proverb 'do', as seen in (4a) where the verb 'did' functions as a predicate substitute for the entire verbal clause. In this sentence, therefore, 'more' will not be considered a clue to anaphoric ellipsis.

(4a) Fourth-graders made significantly more female designations among adult-specified females than did preschoolers.

(4b) Fourth-graders made significantly more female designations among adult-specified females than preschoolers made female designations among adult-specified females.

Quantifier

When used as a quantifer, 'more' means "an additional amount of things, persons, time, etc." and directly precedes the noun phrase it modifies.

'more' + noun phrase

(5) Results show more emphasis on the informational aspects.

Whether the use is anaphoric or not will depend on whether the noun phrase it is modifying is specified any more extensively in prior text.

ANAPHORIC

Numeric Comparison

When a comparison is made between an absolute numeric value and its comparative form (e. g. 'two or more than two'), the text is frequently abbreviated to:

numeral + 'or more' + noun/adjective + noun

This use is a clue to another instance of anaphoric ellipsis in that 'than + numeral' have been ellipsed.

(6a) Ten or more instances of tardiness will result in suspension.

which would be resolved as:

(6b) Ten or more than ten instances of tardiness will result in suspension.

NON-ANAPHORIC

Explicit Standard

When the comparison is being made between some entity and an explicit standard rather than between two clauses, 'more' will directly precede 'than' and be followed by some specific numeric measure. 'More' is never anaphoric in this use.

(7) The average bear weighs more than 2000 pounds.

Intensifier

'More' is used as an intensifier to form the comparative form of both adjectives and adverbs which it premodifies. The adjective or adverb must be of the gradable type, that is, it must be an attribute that may be present to varying degrees. When functioning as an intensifier, 'more' is nonanaphoric. The syntactic environment in which this use of 'more' is found is either:

'more' + gradable adjective

(8) Patients with low MHPG levels are more responsive to treatment with drugs that inhibit norepinephrine uptake.

or

'more' + gradable adverb

(9) Change towards increased assertiveness is more likely to occur when clients realistically assess the possibilities open to them.

The intensifier use occurs only in those sentences in which 'than' does not co-occur with 'more'. Even if 'more' premodifies a gradable adjective, if 'than' is also present, the use of 'more' is to be catagorized as a clausal comparative.

The reason 'more' without 'than' cannot be interpreted as a lexical clue to ellipsis, is that since the writer did not indicate by use of 'than' on what parameter the comparison was to take place, there is more than one interpretation possible and we cannot assume what was intended. For instance in the following piece of text, the comparison is ambiguous because there is no 'than'.

(10) Imagery theory is more of a theory of problem solving and is best examined through the measure of error rate. Linguistic theory is more a measure of sentence processing and is best measured using latencies.

'MOST'

'Most' is an indefinite pronoun of the multal subclass. 'Most' occurred 40 times in the set of 600 abstracts with only 3 of these occurrences being anaphoric for an F.I. of .075.

In that some uses of 'most' were so infrequent as to be singular in their occurrence, rules have not been developed for all observed uses, but rather, in some instances the observed syntactic environment is simply described.

'Most' has four basic functions:

Superlative: 'Most' is used to create the superlative form of both adjectives and adverbs. The meaning of 'most' in such instances is "to the greatest or highest degree". When used to form the superlative of an adjective, one basic syntactic environment would be:

'the' + 'most' + adjective + noun

(1) Short-term instabilities are the most important source of error.

In such a syntactical context, 'most' is non-anaphoric. If, however, the adjective is not followed by a noun:

'the' + 'most' + adjective

the term 'most' is to be considered anaphoric in that it serves as a lexical clue to the ellipsis of the noun.

(2) Six environments were tested for conduciveness to study. Low heat and high light were found to be the most conducive.

Another syntactic environment for 'most' when it forms the superlative of an adjective is basically the same as that used by 'most' to form the superlative of an adverb, and in all instances it is non-anaphoric.

verb 'to be' + 'most' + adjective/adverb

(3) To determine which of several methods was most effective, a series of tests was run.

(4) Short-answer questions are most often inappropriately answered.

Quantifier: 'Most' is used as an indefinite quantifier of mass nouns and plural count nouns, where its meaning is, respectively, "greatest amount of" and "greatest number of". In such uses, 'most' is distinguished from its superlative use, by the fact it is not preceded by 'the'.'Most' is nonanaphoric when it occurs in either of the two most frequent syntactic environments for 'most' as a quantifier:

'most' + noun/adjective + noun

(7) By following the Pritkin diet, most overweight / teenagers lost 10-15 pounds.

or

'most' + 'of'

(8) Most of the students passed the final exam.

However, if the noun which 'most' is quantifying is ellipsed, the use is anaphoric.

'most' + verb

(9) Fifty attendees were bunked together. Most enjoyed the experience.

Noun: 'Most' was observed once in its use as a noun, where the meaning is "the greatest amount", as distinguished from its meaning as a quantifier - "the greatest amount of Its syntax is:

'the' + 'most

(10)Its the most I can get for the car.

Adverb: 'Most' may itself be used as an adverb, not just to form the superlative of an adverb. In such use, 'most' is nonanaphoric and has been observed once in each of the following syntactic environments:

- As an interposing element causing a split infinitive:

'to' + 'most' + verb

(5) The drug was shown to most effect results in premature babies.

- As a displaced adverb:

verb + direct object + 'most'

(6) She baked pies most during the winter months.

'MUCH'

'Much' is another of the indefinite pronouns of the multal subclass. 'Much' occurred a total of 10 times in the 600 abstracts, with 4 of these occurrences being anaphoric for an F.I. of .40. 'Much' was observed in 4 distinct usages. The first use is dependent on prior text to determine whether it is anaphoric or not, while the latter 3 uses are nonanaphoric in all observed instances.

- Clausal comparative

With analysis of 'much' we encounter a 2nd term which frequently serves as a lexical clue to the fact that some words in text have been ellipsed. 'More' is the other term which performed the same function. Both terms are used in comparing two clausal constructions which are semantically parallel. Since the 2nd clause, if fully fleshed out, would be a syntactic duplicate of the first clause, it is common practice to ellide rather than repeat some portion of the common structure.

The syntactic environment in which 'much' functions as this lexical clue to ellipsis is:

'much' + adjective + 'than'

or

'much' + adverb + 'than'

'Than' may or may not immediately follow the adjective or adverb, but the presence of 'than' is essential to indicate that in fact a comparison is being made.

When these particular syntactical environments are encountered in text, it is necessary to check whether the structures of the two clauses are completely parallel or whether some terms have been ellipsed.

(1a) First-borns responded to the anxiety stimulus much differently than later-borns.

(1b) First-borns responded to the anxiety stimulus much differently than later-borns responded to the anxiety stimulus.

In those instances where some text has been ellipsed in the second comparative clause, the use of 'much' will be considered anaphoric, while if no terms have been ellided, the use is non-anaphoric.

The one exception is when 'much' premodifies another term of the class of indefinite pronouns which is itself being used to form the comparative form of the adjective or adverb, e. g.

'much' + 'more' + adjective + 'than'

(2) Ineptitude was much more difficult to pretest for than was disinterest.

When double indefinites occur, the first indefinite is to be thought of as an intensifier and non-anaphoric in all occurrences, while the second indefinite pronoun will be attributed with being the lexical trigger for ellipsis.

- Intensifier

'Much' operates as an intensifier when it precedes an adjective but the clausal hinge 'than' is absent from the construction. In such a use, 'much' is non-anaphoric. The syntactic environment would be:

'much' + adjective + noun

(3) Earlier in his career, Watson had much loftier goals.

- Adjective

'Much' can also function as a simple adjective with the meaning "great in quantity, amount, extent, or degree". Such a use is nonanaphoric and was observed once in the following syntax:

'much' + noun

(4) There is much truth in what you say.

- Noun

'Much' was also oberved once in its nonanaphoric role as a noun in the following context:

verb + 'much' + infinitive clause

(5) His excuse left much to be desired.

PAUCALS

The group of indefinite pronouns known as the paucal subclass, consists of the terms: few, fewer, fewest, little, less, and least.

This group was distributed in the 600 abstracts as follows:

	Psych Abs		Inspec		Total		
	Ana.	Non.	Ana.	Non.	Ana.	Non.	F. I.
few	-	6	-	4	-	10	0
fewer	1	9	-	-	1	9	.10
fewest	-	-	-	-	-	-	-
little	-	4	-	1	-	5	0
less	10	22	-	4	10	26	.28
least	-	8	-	6	-	14	0

A full linguistic analysis , including rule-testing, will be performed only on the term 'less', and the single anaphoric instance of 'fewer' simply described.

'Fewer'

The single anaphoric use of 'fewer' occurred in a sentence composed of two semantically parallel clauses, where the co-occurrence of 'fewer' and 'than' provided the syntactic environment permitting some lexical elements of the second clause to be ellided. Therefore, 'fewer' served as a lexical clue to ellipsis and is attributed with anaphoric status. This is the same basic usage observed with the other two paucal comparatives: 'more' and 'less'.

'LESS'

'Less' is an indefinite pronoun of the paucal subclass. It occurred 36 times in the 600 abstracts with 10 of these uses being anaphoric for an F.I. of .28.

'Less' has 4 specific uses. Two of these uses may be anaphoric or not depending on prior text, while the other two uses are always nonanaphoric.

DEPENDENT ON TEXT

Comparative

When 'less' and 'than' co-occur in a sentence, the environment exists for a comparison to be made between two entities or two clauses. If the comparison is between clauses, the common practice is to ellide some portion of the second clause which is simply a repetition of elements of the first clause. If this type ellipsis occurs, 'less' is to be attributed with being the lexical clue for ellipsis and therefore anaphoric. If, however, the second clause is completely parallel with the first and no words have been ellided, the use of 'less' is nonanaphoric.

The syntax for such a comparative use is:

'less' + ______ + 'than'

where what occurs between the 'less' and the 'than' is highly variable, but the presence of both predicts this usage. When this syntax is encountered, the second clause must be checked for complete syntactic parallelism with the first clause.

(1a) Firstborns reported less death-threat concern than other groups.

which would be resolved as:

(1b) Firstborns reported less death-threat concern than other groups reported death-threat concern.

The one exception to this rule is the idiomatic phrase 'less than' followed by some adjectival form, e.g.:

(2) He was less than honest.

where the true meaning of 'less than' is "by no means". The syntactic environment for this exceptional use is:

'less than' + adjective

If 'less than' is not followed by an adjective, it is to be treated the same as in the clausal comparative usage and the clause that follows 'less than' is to be examined for complete parallel structure with the first clause.

Quantifier

When used as a quantifier, 'less' precedes a noun phrase, which may consist of either:

'less' + noun

or

'less' + adjective + noun

and the term 'than' does not co-occur. In this usage, however, the adjective will not be of the gradable type, which it is in the negative comparative use. Whether the use is anaphoric or not will again depend on whether the noun that 'less' modifies is specified any more extensively in prior text.

(3) Experienced programmers required less warm-up time to score highly.

NON-ANAPHORIC

Negative comparative

'Less' combines with gradable adjectives and adverbs to form their negative comparative form. Gradable refers to an attribute that may be present in varying degrees. In this usage, 'than' never occurs in the construction, which consists of:

'less' + adjective

or

'less' + adverb

(4) Urban lots are considered to be less stable in the current real estate market.

ADVERB

As an adverb, 'less' serves as a "downtoner", lowering the effect of the force of the verb. The syntactic environment for such usage would be:

verb + 'less'

or

verb + direct object + 'less'

In some of its other uses, 'less' may also follow the verb, but in those uses, 'less' would be followed by either an adjective, adverb, noun, or 'than'. When used as an adverb, 'less' is not followed by any of these, and either ends the sentence or is followed by a prepositional phrase.

(5) Students cheated less when dual monitoring devices were used.

'SOME' - group

Of the 'some' group of indefinite pronouns, 'somebody' and 'someone' never occur in the sample set of 600 abstracts, while 'something' only occurs 3 times and is non-anaphoric in each usage. Only 'some' was observed in any anaphoric uses. 'Some' occurred 78 times in the 600 abstracts, with only 2 of these instances being anaphoric for an F.I. of .02.

Non-anaphoric

The major role of 'some' is to serve as a quantifier/ determiner of a noun phrase. In such usage, its meaning is "an unspecified amount or number". 'Some' may immediately precede the noun phrase:

'some' + noun phrase

(1) Some computer-aided design programs are described and illustrated with examples.

or take the of-construction:

'some' + 'of' + noun phrase

(2) Each area is described detailing some of the major proposed solutions to the proposed therein.

Anaphoric

It is possible for the noun phrase which 'some' is serving as determiner for, to be ellipsed. In such a usage, 'some' is anaphoric. The possible syntactic environment for such a use would be either:

'some' + verb phrase

or

'some' + preposition (other than 'of')

(3) The answers were incorrect for a number of reasons. Some were incomplete and some simply wrong.

'ANY' - group

Of the 'any' group of indefinite pronouns, 'anybody' never occurred in the sample set of 600 abstracts. 'Anyone' only occurs twice and 'anything' once. None of these occurrences are anaphoric. 'Any' occurred 24 times, with only 2 on these instances being anaphoric for an F.I. of .08.

ANY

'Any' serves as a quantifier/determiner of a noun phrase and the question of whether the usage is anaphoric or not is answered only by examining prior text to see if the noun that 'any' is modifying is specified earlier in any greater detail.

'EITHER'

'Either' occurred in the 600 abstracts a total of 37 times with 35 of these occurrences being anaphoric for an F.I. of .95. 4 of the 35 anaphoric occurrences were lexical anaphors and the remaining 31 anaphoric uses of 'either' were as lexical clues to ellipsis.

ANAPHORIC

Coordination

The major function of the term 'either' is as an anticipator of a coordinated construction in which the actual coordinator term is 'or'. 'Either - or' may be used to coordinate within phrases or across phrases and clauses, and in both environments 'either' is considered a lexical clue to anaphoric ellipsis.

Phrasal Coordination: The usual syntax for within-phrase coordination is either:

'either' + adjective + 'or' + adjective + noun

(1a) Subjects delivered a prepared speech on either a sexual or a non-sexual topic.

which would be resolved as:

(1b) Subjects delivered a prepared speech on either a sexual topic or subjects delivered a prepared speech on a non-sexual topic.

and perhaps more naturally rephrased as:

(1c) Either subjects delivered a prepared speech on a sexual topic or subjects delivered a prepared speech on a nonsexual topic.

or

form of verb 'to be' + 'either' + adjective + 'or' + adjective

where the attributes expressed by both adjectives are being predicated of the same noun phrase which precedes the verb form of 'to be'.

(2a) Stimuli were either sweet or sour.

which would be resolved as:

(2b) Stimuli were either sweet or stimuli were sour.

and more naturally rephrased as:

(2c) Either stimuli were sweet or stimuli were sour.

Clausal coordination: All other co-occurrences of 'either' and 'or' which do not fit the two syntactic environments described above, will be instances of clausal coordination. Typical use might be:

(3a) The disease either responded paradoxically to treatment or continued to produce severe symptoms.

which would be resolved as:

(3b) The disease either responded paradoxically to treatment or the disease continued to produce symptoms.

and more naturally rephrased as:

(3c) Either the disease responded paradoxically to treatment or the disease continued to produce symptoms.

Determiner

'Either' may function as determiner of a noun phrase and is always anaphoric in such usage. The environment for such use would be the non-occurrence of the term 'or' within the same sentence and the syntax:

'either' + noun phrase

(4) In the second experiment, codeine and demerol were tested. Either drug was found to produce significant side effects.

NON-ANAPHORIC

Nominal

'Either' may function as a nominal, meaning "one or the other". In such a use, 'either' has been nonanaphoric in each occurrence in the test set. When functioning as a nominal, 'either' occurs in a sentence without 'or' and in the following syntax:

'either' + 'of' + noun phrase

(5) Subjects were placed in either of two conditions.

NEGATIVES

There are five negative indefinite pronouns - 'no', 'none', 'nobody', 'nothing' and 'neither'. In the set of 600 abstracts these terms occurred as follows:

'nobody' - no occurrences
'nothing' - one non-anaphoric occurrence
'none' - 3 occurrences; 1 anaphoric, 2 nonanaphoric
'neither' - 3 occurrences; 2 anaphoric, 1 nonanaphoric
'no' - 76 occurrences; 5 anaphoric, 71 nonanaphoric

Rules for 'none' and 'neither' can be easily generated from earlier rule sets written for similarly functioning terms. 'No' is the only negative which occurred sufficiently frequently to warrant a full-scale analysis.

'NONE'

'None' had two distinct uses in the abstracts. The rules governing whether the use was anaphoric or not are the same syntax-matching rules as used for the terms 'most', 'all' and 'each'.

'None' is non-anaphoric in the syntax:

'none' + 'of' + noun phrase

(1) None of the essay questions were responded to in sufficient detail.

'None' serves as a clue to anaphoric ellipsis in the syntax:

'none' + verb form

(2) Three indexing techniques were tested. None improved the results significantly.

'NEITHER'

The two distinct uses of 'neither' were exact syntactic matches to two of the uses that 'either' is put to. Namely, 'neither' is used as a determiner and is anaphoric in the syntax:

'neither' + noun phrase

(3) Subjects were assigned to a control group or the experimental group. Neither group performed exceptionally well.

'Neither' is used as a nominal with the meaning "not one or the other" and is nonanaphoric in the syntax:

'neither'.....+ 'of' + noun phrase

(4) Neither of the fires resulted in any loss of life.

'NO'

'No' occurred a total of 76 times in the 600 abstracts with only 5 of these occurrences being anaphoric for an F.I. of .06.

The one possibly anaphoric use of 'no' is dependent on prior text. The syntax for such use would be:

'no' + noun

where the anaphoric/nonanaphoric decision depends on whether the noun that 'no' is serving as determiner for is specified in any greater detail earlier in text.

(1a) Threshold-raising techniques have been under development for several years. No techniques have yet met the design criteria.

which would be resolved as:

(1b) Threshold-raising techniques have been under development for several years. No threshold-raising techniques have yet met the design criteria.

'No' is always non-anaphoric when premodifying either an adjective:

'no' + adjective + noun

(2) No significant effects were found for birth-status alone.

or an adverb:

'no' + adverb

(3) Physicians believe that quarantine is no longer necessary for victims of tuberculosis.

Residual Adjectives

The 6 remaining P.A.'s that function frequently as adjectives were analyzed with the following results:

	P.A.		INSPEC		TOTALS		
term	Ana.	Non.	Ana.	Non.	Ana.	Non.	F.I.
additional	0	2	1	3	1	5	.16
another	2	3	1	3	3	6	.33
both	18	36	1	25	19	61	.24
else	0	0	0	1	0	1	.00
equal	0	5	0	1	0	6	.00
identical	0	2	0	1	0	3	.00

Since 'else', 'equal', and 'identical' never functioned anaphorically and 'additional' functioned anaphorically only once, no further description of their usage will be presented, nor will they be tested.

Another

'Another' may function in one of three ways:

Non-anaphoric: 'Another' is always non-anaphoric when used to refer to some indeterminate human referent who is presumed to exist but not specified in the text.

(1)Forgiveness of another brings peace of mind.

Dependent-on-text: 'Another' is potentially anaphoric when it serves as modifier in a noun phrase. Whether it is anaphoric or not depends on whether the noun it modifies has been specified in greater detail earlier in text.

(2) There are a variety of ballet styles currently in vogue. One ballet style is the classical and another style is the minimalist.

Anaphoric: 'Another' is always anaphoric when the noun it is intended as modifier for, has been ellipsed.

(3) It has become increasingly difficult to tell one book from another.

Both

'Both' has 2 non-anaphoric uses and 2 anaphoric uses.

Non-anaphoric

The most common use of 'both' is in conjunction with 'and' in what is known as a combinatory coordination. 'Both' is used to stress the inclusion of each of the 2 words or phrases being coordinated. The occurrence of the following syntax always indicates this use:

'both'.......'and'

where the text which separates the 2 terms may be as short as one word or as long as a full phrase.

(4) Both the automaton and its reversal are strongly connected.

or

(5) Constructive assertive alternatives are developed that integrate both the task and feelings.

When 'both' combines with 'of', it again stresses inclusion of each of the items which follow 'of'. 'Both' is always non-anaphoric in such use, although the term or phrase following 'of' is frequently anaphoric.

'both' + 'of'

(6) Both of these techniques have been used in earlier research in content analysis.

Anaphoric

'Both' was observed to function anaphorically in every instance where it served as premodifier in a noun phrase. This use can be recognized by absence of 'and' from the construction and one of the following syntactic patterns:

'both' + noun

(7) Rats and gerbils were tested in the mazes. Both species improved performance following reinforcement trials.

'both' + adjective + noun

(8) Pre-adolescent females and adolescent males were observed in their school settings. Both target groups exhibited self-conscious behavior when advised of the possible observations.

'Both' functions anaphorically when it serves as a pronominal, taking the place of two items referred to earlier in text. In this usage, 'both' occurs wherever a noun might occur and has been observed in the following two patterns:

'both' + verb

(9) Red and yellow were chosen as the stimulus colors. Both elicit similar emotional responses in subjects.

preposition + 'both' (not followed by adjective or noun)

(10) Heavy smokers and frequent drinkers were chosen as subjects. Lack of interest in nutritional concerns has been observed in both.

Adverbs

The linguistic analysis of the 10 adverbs which occur in the set of 600 abstracts has been completed with the following summary results:

	P.A.		Ins.		Total		
	Ana.	Non.	Ana.	Non.	Ana.	Non.	F.I.
here	0	2	1	2	1	4	.20
identically	0	0	0	1	0	1	0
similarly	1	1	0	0	1	1	.50
so	3	8	0	15	3	23	.12
such	16	17	20	27	36	44	.45
then	1	14	1	24	2	38	.05
there	0	36	1	24	1	60	.02
therein	0	0	1	0	1	0	1.00
thus	0	10	0	3	0	13	0
vice versa	1	0	0	0	1	0	1.00
Total	22	88	24	96	46	184	.20

As can be seen by these figures, only 'so' and 'such' demand that rules be written to determine anaphoric from nonanaphoric occurrences. 'Then', with 2 anaphoric occurrences could possibly have rules written, but the single occurrence in each database does not appear to offer any patterned use.

"SO"

'So' occurred a total of 26 times in the 600 abstracts. 15 of these occurrences were in the INSPEC abstracts and all 15 occurences were non-anaphoric. Of the 11 occurrences of 'so' in PSYCH ABS, 3 instances were anaphoric. Total F. I. over 2 databases was 12%. Rules will be written based on the uses of 'so' in just the PSYCH ABS database.

Non-Anaphoric

1. 'So' combines with 'that' to introduce a clause expressing purpose or result. Syntax would be:

   'so' + 'that'

   (1) The velocity of a trolley must be controlled so that the swing of its grab vanishes when the trolley arrrives at a goal position.

2. 'So' combines with foms of the verb 'do' to form a complex anaphoric pro-verb. For this tabulation of anaphoric terms, 'so' in such use will not be counted as anaphoric since in each instance, 'do' has already been credited with anaphoric function. Resolution of the complex pro-verb 'do' reinserts those terms which 'so' substitutes for.

   form of 'do' + 'so'

   (2) Paul has already registered for the new semester and Gene will do so soon.

3. 'So' functions non-anaphorically as an intensifier of either an adjective or adverb, and has the meaning "to a great extent or degree". Recognizable syntax would be:

   'so' + adjective/adverb

   (3) The children were so eager to begin that to wait would have been foolish.

Anaphoric

In all remaining observed instances of use, 'so' functioned as a pro-adverbial. In such uses, its meaning is "such as has been specified or suggested" earlier in text. The contextual syntax of such use was varied in that 'so' can replace an adverb or a whole clause.

(4) They asked whether we were going to the concert. If so, they wanted to go with us.

'SUCH'

'Such' occurred in the set of 600 abstracts a total of 80 times with 36 of these occurrences being anaphoric for an F. I. of 45%. 'Such' has 2 consistently non-anaphoric uses, 2 consistently anaphoric uses, and 1 use dependent on text.

NON-ANAPHORIC

1. 'Such' combines with 'that' to form a compound subordinating conjunction introducing a clause. It is always non-anaphoric in the syntax:

   'such' + 'that'

   (1) The results were presented in a manner such that those unfamiliar with the topic still had no difficulty understanding them.

2. 'Such' combines with 'as' to serve as an explicit indicator that an appositional phrase follows. The appositional phrase provides one or more examples of the noun phrase that precedes it.

   (2) Skills such as providing sympathy, explanation and advice are given.

   in which case the syntax would be:

   noun + 'such' + 'as'

   A possible alternative syntax would be:

   'such' + noun + 'as'

   (3) The basketball teams in contention for first place are such teams as Georgetown, Syracuse and Boston College.

ANAPHORIC

1. 'Such' functions anaphorically as a determiner in a noun phrase and may occur in either:

   'such' + noun

   (4) Tests were administered to students with I . Q.'s bordering on slow learner. Such students frequently presented a problem in placement.

   or

   'such' + adjective + noun

(5) At one time or another, most students take either SAT or GRE tests. Such standard tests are feared by most students.

This use is distinguishable from the second appositional use of 'such' in that 'as' does not follow the noun.

2. 'Such' functions pronominally in the syntax:

'as' + 'such'

(6) The Statue of Liberty is considered by many immigrants to be the symbol of freedom. As such, it was mandatory that the disintegrating structure be restored.

DEPENDENT ON TEXT

'Such' may serve as a predeterminer for an indefinite noun phrase in the syntax:

'such' + 'a'/'an' + noun

(7) System analysts recommended a completely new approach to scheduling deliveries. Such an approach would require extensive groundwork prior to implementation.

Whether the term 'such' is functioning anaphorically or not depends on whether the noun in the phrase has been specified in any greater detail earlier in text.

'Subject'

There are four abbreviated forms of reference to 'Subject' or 'Subjects' in abstracts, namely 'S', 'S's', 'Ss', or 'Ss''. These four possible realization forms were analyzed as a single group, with the following summary results. Abbreviated subject reference occurred in the 300 abstracts from the PsychAbs Database, a total of 213 times with 188 of these occurences being anaphoric for an F. I. of .88. There were no occurrences of any of these 4 abbreviations in the 300 abstracts from INSPEC.

Of the 25 non-anaphoric uses of the Subject abbreviations, 17 are identifiable by matching against 3 possible contextual patterns. The remaining 8 occurrences are much more difficult to tag as non-anaphoric because their syntactic environments are ones in which the same term may be used anaphorically. As a result, it will be necessary to first identify all consistent anaphoric and non-anaphoric patterns of use and then turn to semantic analysis to decide the status of a term occurring in a pattern which can be either anaphoric or non-anaphoric.

The suggested order of pattern-matching will be an interleaving of anaphoric and non-anaphoric rules, rather than first applying all rules of one usage in a sequence followed by all rules of the other usage. The most definite, easily matched patterns will be applied first, with those requiring more complex semantic processing being applied last.

1. Possessive - whenever the two possessive forms are observed, they are anaphoric.

   S's/Ss' + noun

   (1a) 112 college students studied different sets of 16 faces on 3 occasions. Analysis of Ss' consistency showed that more than 50% of them performed consistently.

   (1b) 112 college students studied different sets of 16 faces on 3 occasions. Analysis of 112 college students' consistency showed that more than 50% of them performed consistently.

2. Indefinite quantifier - when terms of this class (e. g. 'each', 'all', 'fewer', 'some', etc.) premodify S/Ss, the S-form was always anaphoric:

   indefinite quantifier + S/Ss

   (2a) Investigated influence of 48 hours of sleep deprivation (SD) in 12 19-30 year old males. Following SD, all Ss showed marked reductions of DNA synthesis.

   (2b) Investigated influence of 48 hours of sleep deprivation (SD) in 12 19-30 year old males. Following SD, all 12

19-30 year old males showed marked reductions of DNA synthesis.

3. Initial introduction of subjects under study is always non-anaphoric and usually of the form:

# + (age) + (adjective) + S/Ss

with either age or descriptive adjective optional, but at least one must be present.

(3a) 8 10 year old female Ss

or

(3b) 8 female Ss

or

(3c) 8 10 year old Ss

4. Another possible pattern for introducing subjects, which again is non-anaphoric, is:

S/Ss + 'were' + description

(4) Ss were nursing home residents with at least 1 year's residency.

5. A further non-anaphoric initial introductory pattern is:

S/Ss + description

(5) 6 Ss, aged 18 to 21, were administered the test.

6. When S/Ss is premodified by a definite article or determiner (e.g. 'the', 'these') the use is anaphoric.

determiner + S/Ss

(6a) Administered the Block Design subtest of the WISC to 550 members of 65 monozygotic twin kinships. Fingerprint ridge counts of the Ss were also analyzed.

(6b) Administered the Block Design subtest of the WISC to 550 members of 65 monozygotic twin kinships. Fingerprint ridge counts of the 550 members of 65 monozygotic kinships were also analyzed.

7. Having identified the above syntactic environments, it appears that the remaining occurrences of S/Ss in the following context will always indicate anaphoric use:

S/Ss + active verb

(7a) Experiment 1 compared recall following semantic orienting instructions, formal orienting instructions, and intentional learning instructions using 19 undergraduate novice chess players. Ss completed the Spatial Visualization Subtest.

(7b) Experiment 1 compared recall following semantic orienting instructions, formal orienting instructions, and intentional learning instructions using 19 undergraduate novice chess players. 19 undergraduate novice chess players completed the Spatial Visualization Subtest.

8. A fairly common syntax for 'S' to occur in, is:

adjective + S/Ss

which could be either anaphoric or non-anaphoric depending on whether the 'S' had been specified formerly. In the greater proportion of cases, the S-form is anaphoric, but it is possible for the S to be referring rather abstractly and generally to subjects without their having been specified earlier.

9. The remaining patterns of use for S/Ss are too singular to permit generalized rule-writing. Therefore, if an occurrence of S/Ss does not match any of the above syntactic patterns, simply check prior text to see if the term has been specified earlier.

APPENDIX D

Test Results of Rule Sets

RESULTS OF TESTS OF RULE SETS

All rules were tested by at least three people. Tester 1 was involved in the project throughout the first year. Tester 2 was not involved except for rule tests. The third and subsequent testers were chosen haphazardly from among students in the School of Information Studies at Syracuse University. The only requirements were that they be native speakers of American English and had not previously tested any other rule sets.

In general, rules were tested by only three people. Whenever one of more did not achieve 90% accuracy, or nearly so, additional people were chosen to test the rules. Exceptions to this practice were when most of the problems arose from rules dealing with whether a concept had been specified earlier in greater detail, e.g. "each", or when the number of examples was so small that one error would drop percentages dramatically.

RESULTS OF TESTS OF RULE SETS

Potential Anaphor	Number Example Tested	Tester 1	Tester 2	Tester 3	Tester 4	Tester 5	Overall	% Error Caused By 2 Rules*
all	48	.792	.875	.812			.879	37.9
another	9	.778	.556	.889			.741	85.7
any	9	1.000	1.000	.889			.963	0.0
both	35	1.000	.970	1.000			.990	0.0
do	50	1.000	.980	.980			.987	0.0
each	74	.889	.824	.849	.864		.857	92.8
either	21	1.000	1.000	1.000			1.000	0.0
it	58	.966	.966	.966			.966	0.0
less	35	.914	.857	.909			.893	27.3
many	26	.846	.769	.731			.782	82.4
more	51	.843	.857	.765	.608		.767	17.0
most	35	.886	.914	.914			.905	85.7
much	10	.900	.900	.700	.889		.846	0.0
neither	3	1.000	.667	1.000			.889	0.0
no	38	.921	.921	.947			.930	87.5
none	3	1.000	1.000	1.000			1.000	0.0
one	44	.977	.886	.977	.795	.866	.900	50.0
ordinals	33	.848	.848	.939			.879	58.3
other	44	.864	.907	.837			.869	76.5
others	16	1.000	.812	.875			.896	0.0
same	38	.921	.842	.789			.851	82.4
so	11	1.000	1.000	1.000			1.000	0.0
some	35	1.000	1.000	.971			.981	0.0
such	61	.869	.918	1.000			.929	38.5
this,that, these,there	42	.881	.810	.833	.857		.845	0.0
which	55	.982	.927	.927	.818	1.000	.932	0.0

*The two rules causing consistent problems dealt with deciding whether a concept was specified in greater detail earlier in text.

APPENDIX E

Retrieval Tests Results

INSPEC Series 100
PsycABS Series 200

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.753857	0.979255	0.320029	
101	1	e	-0.338200	-0.376721	0.938669	0.600638	
101	1	h	-0.084690	-0.080172	0.999317	-0.625587	
101	1	j	-0.605974	-0.522589	0.950945	-1.581806	
101	1	m	-0.742554	-0.731883	0.969913	-0.330174	
101	1	n	-0.692001	-0.694001	0.969060	0.057258	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.698247	0.976141	-0.092007	
101	2	e	-0.339941	-0.375625	0.926487	0.508841	
101	2	f	-0.662621	-0.655125	0.973046	-0.219739	
101	2	g	-0.701503	-0.698217	0.975441	-0.106454	
101	2	h	-0.031633	-0.031112	0.999995	-0.832977	
101	2	j	-0.712239	-0.697572	0.970971	-0.438400	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.722978	0.973538	-0.281127	
101	2	n	-0.741866	-0.750436	0.976085	0.301526	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.047788	0.980452	-0.153916	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.318253	0.999978	-1.128444	
103	1	j	-0.316556	-0.309937	0.986634	-0.208858	
103	1	m	-0.333513	-0.311583	0.979838	-0.564470	
103	1	n	-0.327129	-0.299643	0.983126	-0.769664	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.016685	0.983559	0.350147	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.011232	0.981155	0.545914	
103	2	g	-0.008358	-0.023820	0.982208	0.401651	
103	2	h	-0.314465	-0.313258	0.999984	-1.079661	
103	2	j	-0.350808	-0.351904	0.985081	0.033215	
103	2	l	-0.249678	-0.248764	0.999989	-0.988304	
103	2	m	-0.213815	-0.208176	0.980266	-0.142283	
103	2	n	-0.242461	-0.234154	0.982868	-0.226384	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: [illegible] Anaphoric Class

CENTRAL PRONOUNS

Q	S	TW	Correlation Coefficients r_{ju}	 r_{jr}	 r_{ur}	Significance Level Z	 $p > .05$
104	1	a	-0.071262	-0.071204	0.999[illegible]	-0.018778	
104	1	d	-0.623340	-0.654580	0.977[illegible]	0.833291	
104	1	e	-0.429965	-0.440598	0.990[illegible]	0.389770	
104	1	h	-0.347650	-0.368305	0.996[illegible]	1.180542	
104	1	j	-0.470712	-0.519349	0.985[illegible]	1.430129	
104	1	m	-0.602159	-0.634720	0.98[illegible]	0.951672	
104	1	n	-0.415472	-0.424679	0.987[illegible]	0.288804	
104	2	a	-0.153981	-0.154278	0.999[illegible]	0.080141	
104	2	b	-0.161980	-0.162270	0.999[illegible]	0.165467	
104	2	c	-0.149594	-0.149793	0.999[illegible]	0.059193	
104	2	d	-0.592005	-0.634221	0.970[illegible]	0.960542	
104	2	e	-0.426036	-0.430578	0.986[illegible]	0.150348	
104	2	f	-0.577165	-0.632414	0.966[illegible]	1.156169	
104	2	g	-0.595353	-0.637609	0.966[illegible]	0.936988	
104	2	h	-0.178700	-0.185040	0.999[illegible]	1.439780	
104	2	j	-0.472299	-0.534012	0.983[illegible]	1.663594	
104	2	l	-0.017455	-0.017947	0.999[illegible]	1.327147	
104	2	m	-0.565612	-0.613311	0.976[illegible]	1.173495	
104	2	n	-0.384919	-0.404667	0.986[illegible]	0.628653	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 20[illegible] on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = [illegible]

TW: Term Weighting Schemes: See Result Page R-.

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If the Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.343713	0.991963	-0.623202	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.152646	0.999998	-0.252577	
107	1	j	-0.185056	-0.197007	0.996773	0.624174	
107	1	m	-0.327307	-0.324536	0.993124	-0.103046	
107	1	n	-0.283932	-0.289027	0.997671	0.321043	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.342283	0.990676	-0.461051	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.343860	0.992538	-0.489780	
107	2	g	-0.356079	-0.341662	0.990671	-0.463463	
107	2	h	0.182394	0.181773	0.999988	0.530618	
107	2	j	-0.089937	-0.099660	0.998077	0.649207	
107	2	l	0.074758	0.074224	0.999998	1.003323	
107	2	m	-0.315778	-0.314926	0.993906	-0.033509	
107	2	n	-0.296199	-0.302537	0.997877	0.419752	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.382547	0.986701	0.779711	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.067359	0.999874	0.446446	
109	1	j	-0.131462	-0.124321	0.998326	-0.669802	
109	1	m	-0.328732	-0.342388	0.990582	0.567932	
109	1	n	-0.358384	-0.353675	0.982359	-0.144541	
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.315516	0.986674	1.044622	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.316708	0.988327	0.844025	
109	2	g	-0.289566	-0.318656	0.986581	0.999663	
109	2	h	-0.068850	-0.069123	0.999960	0.165130	
109	2	j	-0.125951	-0.121443	0.999705	-1.006178	
109	2	i	-0.069214	-0.069897	0.998987	0.081812	
109	2	m	-0.296651	-0.305507	0.995931	0.553573	
109	2	n	-0.330370	-0.317918	0.993681	-0.629341	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.601249	0.999626	1.775102	
135	1	d	-0.797236	-0.774166	0.994983	-1.372128	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.047210	0.993752	0.764361	
135	1	j	-0.633972	-0.609244	0.970631	-0.534471	
135	1	m	-0.796876	-0.774092	0.993503	-1.222294	
135	1	n	-0.842132	-0.796913	0.985750	-1.616689	
135	2	a	-0.637630	-0.648643	0.999289	1.447328	
135	2	b	-0.647093	-0.652499	0.999774	1.289618	
135	2	c	-0.635240	-0.644853	0.999490	1.481225	
135	2	d	-0.818316	-0.801744	0.997139	-1.366167	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.798506	0.996762	-1.526497	
135	2	g	-0.816188	-0.797552	0.996860	-1.436400	
135	2	h	-0.001788	-0.002167	0.999999	1.024605	
135	2	j	-0.766710	-0.759376	0.991229	-0.352053	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.800075	0.996594	-1.072395	
135	2	n	-0.826860	-0.791625	0.991970	-1.631944	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.192139	1.000000	0.000000	
142	1	d	-0.255597	-0.235276	0.976352	-0.420026	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.211000	0.780692	-0.452948	
142	1	j	-0.318605	-0.302616	0.825734	-0.124776	
142	1	m	-0.304107	-0.250955	0.939528	-0.693067	
142	1	n	-0.211922	-0.153100	0.922218	-0.661740	
142	2	a	-0.324310	-0.324310	1.000000	0.000000	
142	2	b	-0.435620	-0.435620	1.000000	0.000000	
142	2	c	-0.355980	-0.355980	1.000000	0.000000	
142	2	d	-0.339456	-0.379523	0.949774	0.589922	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.493565	0.947337	0.625258	
142	2	g	-0.352275	-0.394392	0.945752	0.600169	
142	2	h	-0.277419	-0.028591	0.120297	-0.848059	
142	2	j	-0.316247	-0.386914	0.790580	0.511969	
142	2	l	-0.255325	-0.001546	0.395578	-1.038516	
142	2	m	-0.359307	-0.384561	0.924844	0.306575	
142	2	n	-0.254817	-0.264985	0.915232	0.111641	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.251475	0.976040	1.535252	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.300780	0.994599	1.461146	
158	1	j	-0.172847	-0.286021	0.965154	1.631612	
158	1	m	-0.212637	-0.321438	0.971392	1.733685	
158	1	n	-0.190370	-0.307179	0.970019	1.812228	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.182510	0.983876	1.439642	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	-0.045643	0.987622	1.475464	
158	2	g	-0.089306	-0.157537	0.983808	1.427362	
158	2	h	-0.136551	-0.211720	0.988490	1.864984	
158	2	j	-0.103790	-0.200107	0.981001	1.860726	
158	2	l	-0.007262	-0.042342	0.997273	1.778097	
158	2	m	-0.131206	-0.220264	0.981490	1.746647	
158	2	n	-0.120778	-0.217157	0.979994	1.816815	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716651	0.999997	0.091870	
170	1	d	-0.731051	-0.711001	0.997701	-1.642827	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.162283	0.999947	1.852693	
170	1	j	-0.594562	-0.597756	0.995432	0.181383	
170	1	m	-0.636259	-0.627355	0.997374	-0.681088	
170	1	n	-0.515708	-0.521729	0.995990	0.342059	
170	2	a	-0.678385	-0.678507	0.999995	0.231544	
170	2	b	-0.637558	-0.637593	0.999999	0.147925	
170	2	c	-0.678090	-0.678193	0.999996	0.212600	
170	2	d	-0.691708	-0.679235	0.998372	-1.233185	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.688013	0.997958	-0.822354	
170	2	g	-0.689511	-0.677288	0.998387	-1.213704	
170	2	h	0.153761	0.148824	0.999712	0.906317	
170	2	j	-0.553947	-0.568516	0.995968	0.839790	
170	2	i	0.045678	0.042885	0.999911	0.910987	
170	2	m	-0.623943	-0.621827	0.997625	-0.171027	
170	2	n	-0.533728	-0.545733	0.995838	0.674425	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.298091	0.999850	0.167899	
180	1	d	-0.475084	-0.502717	0.995419	1.594363	
180	1	e	-0.484654	-0.481450	0.999093	-0.428602	
180	1	h	-0.234297	-0.250466	0.998063	1.330695	
180	1	j	-0.395090	-0.414369	0.996511	1.243043	
180	1	m	-0.473372	-0.496783	0.995639	1.394231	
180	1	n	-0.428501	-0.440867	0.997166	0.903736	
180	2	a	-0.219548	-0.220523	0.999772	0.233906	
180	2	b	-0.157462	-0.157914	0.999949	0.227314	
180	2	c	-0.213261	-0.214164	0.999834	0.253606	
180	2	d	-0.475058	-0.499907	0.995165	1.406627	
180	2	e	-0.434632	-0.430285	0.998788	-0.488342	
180	2	f	-0.393963	-0.424510	0.994469	1.553497	
180	2	g	-0.468597	-0.494745	0.994704	1.409973	
180	2	h	-0.352302	-0.377496	0.991005	1.001235	
180	2	j	-0.449142	-0.464609	0.998466	1.523503	
180	2	l	-0.126480	-0.139029	0.999041	1.442609	
180	2	m	-0.476596	-0.497468	0.997434	1.606254	
180	2	n	-0.445123	-0.454565	0.998221	0.878111	

NOTES:

Q: Queries 100-199 were searched on INSPEC; 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.227444	0.989033	0.207712	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.141320	0.981922	1.250953	
182	1	j	-0.008062	-0.079609	0.940556	1.101022	
182	1	m	-0.187365	-0.224727	0.953986	0.666024	
182	1	n	-0.166005	-0.205681	0.935865	0.596899	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.231375	0.990351	0.313412	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.148305	0.991905	0.044418	
182	2	g	-0.195747	-0.202798	0.991447	0.291070	
182	2	h	0.188673	0.132517	0.974148	1.322604	
182	2	j	-0.009834	-0.077027	0.923563	0.911687	
182	2	l	0.157129	0.110227	0.979798	1.245173	
182	2	m	-0.198272	-0.220088	0.961063	0.423132	
182	2	n	-0.190478	-0.212590	0.937103	0.337021	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132084	0.999998	-0.626922	
184	1	d	0.129099	0.122673	0.991895	0.227508	
184	1	e	0.064221	0.082308	0.987397	-0.510860	
184	1	h	-0.045491	-0.046270	0.999945	0.333364	
184	1	j	0.003093	0.003235	0.997638	-0.009207	
184	1	m	0.080914	0.083962	0.994253	-0.127560	
184	1	n	-0.007330	0.006432	0.996203	-0.706219	
184	2	a	-0.148937	-0.148704	0.999998	-0.604388	
184	2	b	-0.139155	-0.139063	1.000000	-0.641365	
184	2	c	-0.145872	-0.145648	0.999999	-0.613183	
184	2	d	0.128770	0.131270	0.997377	-0.155701	
184	2	e	0.142193	0.174389	0.985549	-0.856993	
184	2	f	0.039625	0.039512	0.998732	0.010048	
184	2	g	0.121064	0.123144	0.997688	-0.137803	
184	2	h	-0.032087	-0.034502	0.999285	0.285808	
184	2	j	0.038815	0.045164	0.998193	-0.472689	
184	2	l	-0.004347	-0.004693	0.999980	0.244317	
184	2	m	0.076227	0.083662	0.997895	-0.514018	
184	2	n	-0.024803	-0.011222	0.997728	-0.901239	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
			Correlation Coefficients			**Significance Level**	
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.743167	0.991340	-0.131274	
101	1	e	-0.338200	-0.420836	0.919309	1.128781	
101	1	h	-0.084690	-0.073091	0.999172	-1.456540	
101	1	j	-0.605974	-0.549778	0.974467	-1.492881	
101	1	m	-0.742554	-0.733313	0.987130	-0.434329	
101	1	n	-0.692001	-0.690408	0.979511	-0.055786	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.696190	0.970633	-0.143759	
101	2	e	-0.339941	-0.424059	0.937578	1.302068	
101	2	f	-0.662621	-0.639035	0.958457	-0.549887	
101	2	g	-0.701503	-0.693389	0.967768	-0.228818	
101	2	h	-0.031633	-0.031112	0.999696	-0.107922	
101	2	j	-0.712239	-0.694885	0.968140	-0.493827	
101	2	l	-0.001532	-0.001634	1.000000	0.527291	
101	2	m	-0.731639	-0.721823	0.968284	-0.291164	
101	2	n	-0.741866	-0.741629	0.972862	-0.007803	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.045412	0.968317	-0.167196	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.327984	0.996000	0.478664	
103	1	j	-0.316556	-0.347186	0.966486	0.613698	
103	1	m	-0.333513	-0.291407	0.960713	-0.773208	
103	1	n	-0.327129	-0.299074	0.972824	-0.619945	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.000189	0.969297	-0.069928	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.031633	0.967820	0.023628	
103	2	g	-0.008358	-0.002717	0.967267	-0.108007	
103	2	h	-0.314465	-0.322391	0.998423	0.726812	
103	2	j	-0.350808	-0.355834	0.970815	0.109041	
103	2	l	-0.249678	-0.247393	0.999388	-0.330210	
103	2	m	-0.213815	-0.180145	0.962082	-0.610890	
103	2	n	-0.242461	-0.214744	0.972087	-0.590057	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.696523	0.961874	1.476849	
104	1	e	-0.429965	-0.442825	0.994226	0.592072	
104	1	h	-0.347650	-0.325655	0.993938	-0.941912	
104	1	j	-0.470712	-0.547544	0.969694	1.552337	
104	1	m	-0.602159	-0.665474	0.966753	1.350265	
104	1	n	-0.415472	-0.439667	0.973957	0.523398	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.601739	0.960396	0.193667	
104	2	e	-0.426036	-0.444296	0.990461	0.653143	
104	2	f	-0.577165	-0.553011	0.943020	-0.389168	
104	2	g	-0.595353	-0.594504	0.955413	-0.015916	
104	2	h	-0.178700	-0.206653	0.990873	0.940980	
104	2	j	-0.472299	-0.515531	0.971369	0.917733	
104	2	l	-0.017455	-0.022193	0.999858	1.256017	
104	2	m	-0.565612	-0.585049	0.966321	0.409573	
104	2	n	-0.384919	-0.407576	0.981547	0.572858	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.364412	0.995629	0.127671	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.150710	0.999975	1.080306	
107	1	j	-0.185056	-0.202360	0.998655	1.393102	
107	1	m	-0.327307	-0.340325	0.997626	0.820948	
107	1	n	-0.283932	-0.311586	0.996012	1.319608	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.366343	0.996669	0.525359	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.366301	0.995186	0.396901	
107	2	g	-0.356079	-0.366054	0.996413	0.519198	
107	2	h	0.182394	0.179168	0.999933	1.161495	
107	2	j	-0.089937	-0.113065	0.997489	1.350585	
107	2	l	0.074758	0.072582	0.999973	1.221262	
107	2	m	-0.315778	-0.329978	0.997207	0.822671	
107	2	n	-0.296199	-0.317888	0.996274	1.078252	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220391	0.999993	-0.184291	
109	1	d	-0.360544	-0.326320	0.990670	-1.419951	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.064229	0.999978	-1.471942	
109	1	j	-0.131462	-0.107257	0.998547	-2.425419	(****
109	1	m	-0.328732	-0.284495	0.990214	-1.763621	
109	1	n	-0.358384	-0.310506	0.983461	-1.487103	
109	2	a	-0.228134	-0.227988	0.999989	-0.169656	
109	2	b	-0.233510	-0.233440	0.999998	-0.187991	
109	2	c	-0.234593	-0.234471	0.999992	-0.165720	
109	2	d	-0.285177	-0.245864	0.978083	-1.045704	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.251748	0.977819	-1.114459	
109	2	g	-0.289566	-0.248342	0.977525	-1.083568	
109	2	h	-0.068850	-0.066723	0.999987	-2.218802	(****
109	2	j	-0.125951	-0.109958	0.998929	-1.869873	
109	2	l	-0.069214	-0.067864	0.998908	-0.155961	
109	2	m	-0.296651	-0.242497	0.977300	-1.412894	
109	2	n	-0.330370	-0.263466	0.971533	-1.567337	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.811397	0.996153	1.049544	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.038609	0.988267	0.326041	
135	1	j	-0.633972	-0.651346	0.971378	0.390669	
135	1	m	-0.796876	-0.812029	0.995231	1.014222	
135	1	n	-0.842132	-0.816545	0.986291	-1.067418	
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.826965	0.997382	0.831461	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.827055	0.994218	0.538514	
135	2	g	-0.816188	-0.825651	0.996668	0.805419	
135	2	h	-0.001788	-0.002294	0.999996	0.755600	
135	2	j	-0.766710	-0.782876	0.990348	0.739402	
135	2	l	-0.001329	-0.001408	1.000000	0.465540	
135	2	m	-0.813643	-0.823161	0.996399	0.776941	
135	2	n	-0.826860	-0.799486	0.988903	-1.195659	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.309716	0.982386	1.298286	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.243869	0.895964	-0.333009	
142	1	j	-0.318605	-0.343277	0.877584	0.231147	
142	1	m	-0.304107	-0.308791	0.956178	0.072539	
142	1	n	-0.211922	-0.211875	0.977826	-0.001009	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.400899	0.965546	1.087382	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.497603	0.958492	0.771049	
142	2	g	-0.352275	-0.412972	0.962126	1.030924	
142	2	h	-0.277419	-0.076194	0.295975	-0.764884	
142	2	j	-0.316247	-0.376493	0.927635	0.734979	
142	2	l	-0.255325	-0.002870	0.400343	-1.037058	
142	2	m	-0.359307	-0.378244	0.990437	0.639633	
142	2	n	-0.254817	-0.273608	0.995289	0.871430	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.189500	0.991659	0.781175	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.266503	0.999645	0.778598	
158	1	j	-0.172847	-0.211844	0.982066	0.783766	
158	1	m	-0.212637	-0.236190	0.991949	0.710944	
158	1	n	-0.190370	-0.196723	0.988442	0.159367	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.173253	0.986406	1.355710	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	-0.073108	0.977403	1.578636	
158	2	g	-0.089306	-0.158132	0.983367	1.420683	
158	2	h	-0.136551	-0.153608	0.997716	0.952612	
158	2	j	-0.103790	-0.140384	0.994449	1.306359	
158	2	l	-0.007262	-0.013926	0.999693	1.006950	
158	2	m	-0.131206	-0.181133	0.990884	1.394890	
158	2	n	-0.120778	-0.138402	0.994413	0.628804	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.685706	0.992263	-1.401858	
170	1	d	-0.731051	-0.646183	0.983124	-2.223138	<****
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.173152	0.999647	-1.085005	
170	1	j	-0.594562	-0.423001	0.950077	-2.435965	<****
170	1	m	-0.636259	-0.508180	0.968596	-2.331625	<****
170	1	n	-0.515708	-0.367760	0.960695	-2.360772	<****
170	2	a	-0.678385	-0.657053	0.994734	-1.159620	
170	2	b	-0.637558	-0.621966	0.996295	-0.985777	
170	2	c	-0.678090	-0.656719	0.994744	-1.162284	
170	2	d	-0.691708	-0.616322	0.990087	-2.414649	<****
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.603192	0.984067	-2.389533	<****
170	2	g	-0.689511	-0.609665	0.989588	-2.462709	<****
170	2	h	0.153761	0.168288	0.984573	-0.365236	
170	2	j	-0.553947	-0.397350	0.967510	-2.664149	<****
170	2	l	0.045678	0.080569	0.978677	-0.738145	
170	2	m	-0.623943	-0.508156	0.978974	-2.494178	<****
170	2	n	-0.533728	-0.396867	0.974306	-2.624840	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.299409	0.999723	0.416233	
180	1	d	-0.475084	-0.443615	0.984269	-0.987770	
180	1	e	-0.484654	-0.481425	0.961570	-0.066709	
180	1	h	-0.234297	-0.249766	0.998054	1.270601	
180	1	j	-0.395090	-0.381667	0.987670	-0.463165	
180	1	m	-0.473372	-0.448727	0.984676	-0.787945	
180	1	n	-0.428501	-0.413271	0.990886	-0.619616	
180	2	a	-0.219548	-0.222590	0.999568	0.530523	
180	2	b	-0.157462	-0.159036	0.999904	0.573887	
180	2	c	-0.213261	-0.215989	0.999684	0.554789	
180	2	d	-0.475058	-0.417685	0.972066	-1.331525	
180	2	e	-0.434632	-0.425915	0.944135	-0.144878	
180	2	f	-0.393963	-0.318482	0.965940	-1.525360	
180	2	g	-0.468597	-0.406472	0.969999	-1.383768	
180	2	h	-0.352302	-0.376805	0.990987	0.973103	
180	2	j	-0.449142	-0.430976	0.990650	-0.735529	
180	2	l	-0.126480	-0.138604	0.999041	1.394080	
180	2	m	-0.476596	-0.442289	0.983133	-1.038756	
180	2	n	-0.445123	-0.427303	0.992705	-0.813962	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192258	0.999981	-1.848643	
182	1	d	-0.221779	-0.203140	0.977838	-0.479331	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.150622	0.970196	0.771710	
182	1	j	-0.008062	0.012438	0.893354	-0.234918	
182	1	m	-0.187365	-0.151624	0.932188	-0.521376	
182	1	n	-0.166005	-0.126694	0.920920	-0.529116	
182	2	a	-0.174304	-0.171877	0.999975	-1.851786	
182	2	b	-0.101067	-0.100460	0.999999	-1.865656	
182	2	c	-0.137425	-0.135802	0.999989	-1.884402	
182	2	d	-0.223362	-0.213458	0.985832	-0.319006	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.146636	0.992152	-0.026167	
182	2	g	-0.195747	-0.189982	0.988383	-0.203983	
182	2	h	0.188673	0.159003	0.964936	0.602036	
182	2	j	-0.009834	0.004380	0.917764	-0.185477	
182	2	l	0.157129	0.155419	0.947313	0.028239	
182	2	m	-0.198272	-0.185293	0.962634	-0.256040	
182	2	n	-0.190478	-0.174692	0.954375	-0.281339	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.045449	0.969904	1.532534	
184	1	e	0.064221	-0.020261	0.958485	1.314819	
184	1	h	-0.045491	-0.045720	0.999955	0.108175	
184	1	j	0.003093	-0.040827	0.982528	1.051650	
184	1	m	0.080914	-0.008729	0.968523	1.603064	
184	1	n	-0.007330	-0.073038	0.984191	1.655927	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.037447	0.945520	1.244465	
184	2	e	0.142193	0.014647	0.914287	1.389034	
184	2	f	0.039625	-0.028108	0.949411	0.953915	
184	2	g	0.121064	0.030929	0.944114	1.212131	
184	2	h	-0.032087	-0.034327	0.999309	0.269516	
184	2	j	0.038815	-0.031108	0.963138	1.153730	
184	2	l	-0.004347	-0.004624	0.999980	0.195383	
184	2	m	0.076227	-0.011179	0.951593	1.260306	
184	2	n	-0.024803	-0.089227	0.978990	1.409272	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.749064	0.983325	0.153600	
101	1	e	-0.338200	-0.317505	0.989662	-0.774259	
101	1	h	-0.084690	-0.079397	0.999559	-0.911198	
101	1	j	-0.605974	-0.560081	0.973645	-1.222811	
101	1	m	-0.742554	-0.734479	0.979690	-0.304004	
101	1	n	-0.692001	-0.693713	0.980652	0.061782	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.696451	0.993459	-0.286243	
101	2	e	-0.339941	-0.327153	0.994202	-0.640832	
101	2	f	-0.662621	-0.667009	0.993487	0.262321	
101	2	g	-0.701503	-0.696718	0.993454	-0.297964	
101	2	h	-0.031633	-0.031835	0.999998	0.505168	
101	2	j	-0.712239	-0.700850	0.992115	-0.646878	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.723385	0.993237	-0.523996	
101	2	n	-0.741866	-0.741049	0.993715	-0.055436	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.187604	0.944741	1.989412	<****
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.325822	0.994913	0.313763	
103	1	j	-0.316556	-0.387584	0.981964	1.906665	
103	1	m	-0.333513	-0.465447	0.949744	2.140543	<****
103	1	n	-0.327129	-0.444652	0.958036	2.081287	<****
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.111776	0.962889	1.952819	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	-0.071671	0.969418	2.078721	<****
103	2	g	-0.008358	-0.117750	0.963068	1.982336	<****
103	2	h	-0.314465	-0.325404	0.990424	0.408408	
103	2	j	-0.350808	-0.408677	0.977241	1.413670	
103	2	l	-0.249678	-0.266875	0.993847	0.784268	
103	2	m	-0.213815	-0.328665	0.959766	2.033012	<****
103	2	n	-0.242461	-0.354892	0.963925	2.106945	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.627595	0.982378	0.130218	
104	1	e	-0.429965	-0.364141	0.968826	-1.264247	
104	1	h	-0.347650	-0.356833	0.999429	1.278629	
104	1	j	-0.470712	-0.494600	0.997058	1.524916	
104	1	m	-0.602159	-0.610769	0.984256	0.272978	
104	1	n	-0.415472	-0.376727	0.980639	-0.950236	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.577360	0.981243	-0.416033	
104	2	e	-0.426036	-0.347535	0.961916	-1.356476	
104	2	f	-0.577165	-0.550421	0.980798	-0.731410	
104	2	g	-0.595353	-0.578368	0.980448	-0.472944	
104	2	h	-0.178700	-0.173816	0.999808	-1.128083	
104	2	j	-0.472299	-0.464022	0.989392	-0.287427	
104	2	l	-0.017455	-0.016772	0.999998	-1.475195	
104	2	m	-0.565612	-0.548504	0.980712	-0.467734	
104	2	n	-0.384919	-0.325057	0.978652	-1.364343	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.372445	0.986546	0.289865	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.153752	0.999995	-1.614661	
107	1	j	-0.185056	-0.168829	0.997112	-0.892639	
107	1	m	-0.327307	-0.331444	0.992286	0.145449	
107	1	n	-0.283932	-0.273002	0.995338	-0.485199	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.360105	0.977705	0.072999	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.349129	0.980249	-0.185285	
107	2	g	-0.356079	-0.357375	0.977327	0.026876	
107	2	h	0.182394	0.182384	0.999985	0.008059	
107	2	j	-0.089937	-0.095323	0.997719	0.330179	
107	2	l	0.074758	0.074475	0.999997	0.514123	
107	2	m	-0.315778	-0.330489	0.985489	0.375926	
107	2	n	-0.296199	-0.310458	0.993636	0.545675	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.368055	0.995997	0.484636	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.067203	0.999885	0.412155	
109	1	j	-0.131462	-0.115287	0.998461	-1.579334	
109	1	m	-0.328732	-0.313539	0.995504	-0.907232	
109	1	n	-0.358384	-0.327633	0.992719	-1.443222	
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.295662	0.996742	0.729392	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.303509	0.997940	0.847540	
109	2	g	-0.289566	-0.299859	0.997042	0.752365	
109	2	h	-0.068850	-0.069299	0.999962	0.276075	
109	2	j	-0.125951	-0.117146	0.999679	-1.881577	
109	2	l	-0.069214	-0.070093	0.998916	0.101836	
109	2	m	-0.296651	-0.287690	0.997368	-0.694039	
109	2	n	-0.330370	-0.304761	0.995856	-1.577314	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.799843	0.995308	0.184390	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.036135	0.998851	0.828982	
135	1	j	-0.633972	-0.637349	0.973348	0.078466	
135	1	m	-0.796876	-0.802317	0.994092	0.342309	
135	1	n	-0.842132	-0.818151	0.994338	-1.444957	
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.814195	0.996665	-0.357978	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.810473	0.996826	-0.734329	
135	2	g	-0.816188	-0.811116	0.996606	-0.432548	
135	2	h	-0.001788	-0.002041	0.999999	0.916607	
135	2	j	-0.766710	-0.766538	0.988823	-0.007431	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.810744	0.995716	-0.221203	
135	2	n	-0.826860	-0.794102	0.993203	-1.645399	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.222814	0.996263	-1.681147	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.209615	0.780640	-0.462259	
142	1	j	-0.318605	-0.272716	0.815363	-0.346104	
142	1	m	-0.304107	-0.208429	0.939471	-1.235898	
142	1	n	-0.211922	-0.088551	0.928968	-1.446407	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.320403	0.993939	-0.794982	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.435825	0.994012	-0.733332	
142	2	g	-0.352275	-0.331839	0.993523	-0.827976	
142	2	h	-0.277419	-0.027038	0.120442	-0.853484	
142	2	j	-0.316247	-0.317895	0.788080	0.011700	
142	2	l	-0.255325	-0.001435	0.395597	-1.038994	
142	2	m	-0.359307	-0.298909	0.946984	-0.853515	
142	2	n	-0.254817	-0.173665	0.935669	-1.009203	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.118249	0.963858	-0.627849	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.259392	0.991015	-0.050337	
158	1	j	-0.172847	-0.128728	0.950448	-0.530626	
158	1	m	-0.212637	-0.156002	0.948555	-0.672173	
158	1	n	-0.190370	-0.140259	0.940936	-0.553406	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.080389	0.965574	-0.481211	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.039111	0.971860	-0.359773	
158	2	g	-0.089306	-0.057415	0.964356	-0.448257	
158	2	h	-0.136551	-0.155115	0.988556	0.463965	
158	2	j	-0.103790	-0.074783	0.960184	-0.386262	
158	2	l	-0.007262	-0.012984	0.997264	0.289458	
158	2	m	-0.131206	-0.092580	0.953584	-0.477538	
158	2	n	-0.120778	-0.083182	0.952300	-0.458020	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	d	-0.731051	-0.710474	0.987554	-0.801223	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.162578	0.999901	1.260646	
170	1	j	-0.594562	-0.606288	0.958230	0.222149	
170	1	m	-0.636259	-0.628139	0.984468	-0.259331	
170	1	n	-0.515708	-0.513102	0.977975	-0.063236	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.681200	0.994732	-0.605948	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.688632	0.996515	-0.593786	
170	2	g	-0.689511	-0.678255	0.995194	-0.675007	
170	2	h	0.153761	0.149720	0.999621	0.646697	
170	2	j	-0.553947	-0.561281	0.988401	0.252795	
170	2	l	0.045678	0.042893	0.999890	0.817551	
170	2	m	-0.623943	-0.618165	0.993232	-0.275853	
170	2	n	-0.533728	-0.531134	0.990012	-0.094582	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.299409	0.999723	0.416233	
180	1	d	-0.475084	-0.453109	0.993410	-1.065212	
180	1	e	-0.484654	-0.486419	0.997938	0.157082	
180	1	h	-0.234297	-0.237410	0.999924	1.293968	
180	1	j	-0.395090	-0.371335	0.993344	-1.103068	
180	1	m	-0.473372	-0.453674	0.995890	-1.203442	
180	1	n	-0.428501	-0.418932	0.998212	-0.876201	
180	2	a	-0.219548	-0.222590	0.999568	0.530523	
180	2	b	-0.157462	-0.159036	0.999904	0.573887	
180	2	c	-0.213261	-0.215989	0.999684	0.554789	
180	2	d	-0.475058	-0.445913	0.990742	-1.185910	
180	2	e	-0.434632	-0.429700	0.989951	-0.192910	
180	2	f	-0.393963	-0.367201	0.992613	-1.176779	
180	2	g	-0.468597	-0.438665	0.990722	-1.211435	
180	2	h	-0.352302	-0.363534	0.998512	1.092303	
180	2	j	-0.449142	-0.436498	0.996630	-0.851719	
180	2	l	-0.126480	-0.131559	0.999839	1.424329	
180	2	m	-0.476596	-0.461856	0.996425	-0.974963	
180	2	n	-0.445123	-0.440131	0.999104	-0.654365	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.219681	0.995821	-0.124511	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.183517	0.999973	1.602480	
182	1	j	-0.008062	-0.024828	0.993841	0.799560	
182	1	m	-0.187365	-0.203022	0.993023	0.714507	
182	1	n	-0.166005	-0.197479	0.991731	1.313595	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.207288	0.997041	-1.128704	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.137067	0.998372	-0.952933	
182	2	g	-0.195747	-0.181955	0.997421	-1.032913	
182	2	h	0.188673	0.183247	0.999847	1.660796	
182	2	j	-0.009834	-0.033027	0.995221	1.255797	
182	2	l	0.157129	0.154619	0.999964	1.570127	
182	2	m	-0.198272	-0.201929	0.996056	0.222398	
182	2	n	-0.190478	-0.214016	0.994976	1.264654	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.173147	0.988817	-1.329493	
184	1	e	0.064221	0.081639	0.977926	-0.371750	
184	1	h	-0.045491	-0.045826	0.999957	0.161818	
184	1	j	0.003093	0.009654	0.984701	-0.167749	
184	1	m	0.080914	0.115369	0.991236	-1.168877	
184	1	n	-0.007330	0.008813	0.993241	-0.620961	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.159767	0.992321	-1.128636	
184	2	e	0.142193	0.155850	0.988994	-0.416260	
184	2	f	0.039625	0.066788	0.992792	-1.013214	
184	2	g	0.121064	0.152444	0.992028	-1.120426	
184	2	h	-0.032087	-0.034308	0.999341	0.273748	
184	2	j	0.038815	0.057438	0.998615	-1.583726	
184	2	l	-0.004347	-0.004693	0.999980	0.244317	
184	2	m	0.076227	0.103527	0.996773	-1.524129	
184	2	n	-0.024803	-0.015615	0.998805	-0.840834	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.749496	0.998202	0.516943	
101	1	e	-0.338200	-0.302868	0.984463	-1.072558	
101	1	h	-0.084690	-0.079805	0.999592	-0.874708	
101	1	j	-0.605974	-0.557030	0.986246	-1.742009	
101	1	m	-0.742554	-0.742365	0.997346	-0.019766	
101	1	n	-0.692001	-0.695962	0.996262	0.323711	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.697187	0.998626	-0.521328	
101	2	e	-0.339941	-0.317809	0.993168	-1.016145	
101	2	f	-0.662621	-0.659921	0.998299	-0.314047	
101	2	g	-0.701503	-0.697474	0.998507	-0.522256	
101	2	h	-0.031633	-0.031691	0.999999	0.182624	
101	2	j	-0.712239	-0.702921	0.997943	-1.016993	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.726186	0.998407	-0.708641	
101	2	n	-0.741866	-0.739675	0.998179	-0.274987	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.052729	0.999982	-1.023534	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.320317	0.999982	0.527357	
103	1	j	-0.316556	-0.326555	0.996549	0.621214	
103	1	m	-0.333513	-0.329265	0.999565	-0.744875	
103	1	n	-0.327129	-0.322722	0.999634	-0.839744	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.004200	0.999985	0.421409	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.032626	0.999996	0.423198	
103	2	g	-0.008358	-0.008768	0.999987	0.391681	
103	2	h	-0.314465	-0.315672	0.999963	0.718655	
103	2	j	-0.350808	-0.356665	0.998460	0.551479	
103	2	l	-0.249678	-0.249504	0.999992	-0.220220	
103	2	m	-0.213815	-0.213963	0.999843	0.041967	
103	2	n	-0.242461	-0.243094	0.999793	0.157330	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.619571	0.997626	-0.311435	
104	1	e	-0.429965	-0.421939	0.998045	-0.631066	
104	1	h	-0.347650	-0.337124	0.994212	-0.464893	
104	1	j	-0.470712	-0.447830	0.993779	-1.016407	
104	1	m	-0.602159	-0.595621	0.996509	-0.434985	
104	1	n	-0.415472	-0.368309	0.986257	-1.355409	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.566358	0.996800	-1.637171	
104	2	e	-0.426036	-0.409160	0.991247	-0.624896	
104	2	f	-0.577165	-0.537765	0.994746	-1.892411	
104	2	g	-0.595353	-0.566049	0.996284	-1.723080	
104	2	h	-0.178700	-0.147296	0.995153	-1.440278	
104	2	j	-0.472299	-0.453364	0.995866	-1.032890	
104	2	l	-0.017455	-0.014014	0.999954	-1.610789	
104	2	m	-0.565612	-0.543397	0.996573	-1.378252	
104	2	n	-0.384919	-0.342904	0.991613	-1.522880	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.348075	0.997812	-0.901149	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.152817	0.999998	-0.678450	
107	1	j	-0.185056	-0.180054	0.999474	-0.645763	
107	1	m	-0.327307	-0.319678	0.998527	-0.610309	
107	1	n	-0.283932	-0.282114	0.999427	-0.230839	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.340723	0.996884	-0.878567	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.340285	0.996814	-0.938955	
107	2	g	-0.356079	-0.339461	0.996661	-0.887019	
107	2	h	0.182394	0.181934	0.999988	0.394991	
107	2	j	-0.089937	-0.085541	0.999382	-0.517402	
107	2	l	0.074758	0.074394	0.999998	0.696253	
107	2	m	-0.315778	-0.306067	0.997986	-0.661291	
107	2	n	-0.296199	-0.292580	0.999205	-0.390968	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.375424	0.997808	1.286514	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.067629	0.999875	0.539338	
109	1	j	-0.131462	-0.119429	0.998371	-1.143141	
109	1	m	-0.328732	-0.328888	0.996127	0.010069	
109	1	n	-0.358384	-0.353895	0.992131	-0.206230	
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.298763	0.997576	1.093045	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.303923	0.998215	0.948744	
109	2	g	-0.289566	-0.302522	0.997656	1.061412	
109	2	h	-0.068850	-0.069777	0.999947	0.483698	
109	2	j	-0.125951	-0.120007	0.999550	-1.073770	
109	2	l	-0.069214	-0.070290	0.998914	0.124573	
109	2	m	-0.296651	-0.296048	0.996755	-0.042274	
109	2	n	-0.330370	-0.322610	0.993413	-0.384951	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.798253	0.999966	0.806227	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.035626	0.998908	0.805472	
135	1	j	-0.633972	-0.644212	0.980281	0.276614	
135	1	m	-0.796876	-0.800067	0.999716	0.876231	
135	1	n	-0.842132	-0.811076	0.995357	-1.829803	
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.819110	0.999765	0.261981	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.818926	0.999918	-0.019273	
135	2	g	-0.816188	-0.817176	0.999821	0.370452	
135	2	h	-0.001788	-0.002120	0.999999	1.315987	
135	2	j	-0.766710	-0.778213	0.995618	0.773603	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.816604	0.999662	0.781532	
135	2	n	-0.826860	-0.797563	0.995690	-1.778178	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.296225	0.979321	0.902495	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.249532	0.915207	-0.306799	
142	1	j	-0.318605	-0.306749	0.918706	-0.135216	
142	1	m	-0.304107	-0.288012	0.995506	-0.771676	
142	1	n	-0.211922	-0.180804	0.992868	-1.154098	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.373569	0.976897	0.737341	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.472702	0.979268	0.487727	
142	2	g	-0.352275	-0.385374	0.976469	0.712576	
142	2	h	-0.277419	-0.054444	0.218294	-0.805042	
142	2	j	-0.316247	-0.345299	0.916169	0.328365	
142	2	l	-0.255325	-0.002208	0.398154	-1.037952	
142	2	m	-0.359307	-0.348580	0.998619	-0.943022	
142	2	n	-0.254817	-0.238696	0.995236	-0.741030	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.167073	0.999889	1.058125	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.268308	0.999643	1.034222	
158	1	j	-0.172847	-0.187003	0.991193	0.405551	
158	1	m	-0.212637	-0.221988	0.999772	1.656869	
158	1	n	-0.190370	-0.210136	0.992971	0.635701	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.115931	0.999840	0.413479	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.014657	0.999962	0.711746	
158	2	g	-0.089306	-0.091100	0.999877	0.429229	
158	2	h	-0.136551	-0.175037	0.997409	2.004835	<****
158	2	j	-0.103790	-0.120490	0.998705	1.233151	
158	2	l	-0.007262	-0.020390	0.999677	1.932696	
158	2	m	-0.131206	-0.139479	0.999815	1.617782	
158	2	n	-0.120778	-0.136954	0.996159	0.695833	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	d	-0.731051	-0.734416	0.998726	0.423922	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.165178	0.999969	0.810368	
170	1	j	-0.594562	-0.600246	0.997641	0.447377	
170	1	m	-0.636259	-0.640931	0.998552	0.487940	
170	1	n	-0.515708	-0.523314	0.998513	0.704176	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.696222	0.999350	0.742064	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.699257	0.999797	0.687098	
170	2	g	-0.689511	-0.694292	0.999377	0.798120	
170	2	h	0.153761	0.153183	0.999984	0.454847	
170	2	j	-0.553947	-0.558604	0.999150	0.587486	
170	2	l	0.045678	0.045931	0.999997	-0.476418	
170	2	m	-0.623943	-0.628105	0.999382	0.652716	
170	2	n	-0.533728	-0.539636	0.999363	0.841414	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.301480	0.999714	0.861202	
180	1	d	-0.475084	-0.475553	0.994769	0.026111	
180	1	e	-0.484654	-0.484153	0.994886	-0.028340	
180	1	h	-0.234297	-0.225405	0.993965	-0.415622	
180	1	j	-0.395090	-0.389419	0.997850	-0.469120	
180	1	m	-0.473372	-0.459382	0.993912	-0.712245	
180	1	n	-0.428501	-0.410778	0.996425	-1.138991	
180	2	a	-0.219548	-0.225084	0.999557	0.950939	
180	2	b	-0.157462	-0.160095	0.999902	0.951170	
180	2	c	-0.213261	-0.218094	0.999676	0.969378	
180	2	d	-0.475058	-0.491692	0.996970	1.196160	
180	2	e	-0.434632	-0.427009	0.987359	-0.265615	
180	2	f	-0.393963	-0.410412	0.999072	2.002848	<****
180	2	g	-0.468597	-0.486036	0.997586	1.390247	
180	2	h	-0.352302	-0.339907	0.967043	-0.257485	
180	2	j	-0.449142	-0.451712	0.998800	0.293474	
180	2	l	-0.126480	-0.119456	0.996554	-0.426230	
180	2	m	-0.476596	-0.482945	0.997654	0.526265	
180	2	n	-0.445123	-0.441612	0.998257	-0.331234	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.205365	0.995631	-0.949108	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.189397	0.999911	-1.475952	
182	1	j	-0.008062	0.023412	0.994536	-1.593812	
182	1	m	-0.187365	-0.149671	0.990448	-1.460139	
182	1	n	-0.166005	-0.120033	0.987143	-1.530116	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.214499	0.997090	-0.629463	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.143760	0.999023	-0.422065	
182	2	g	-0.195747	-0.188411	0.997777	-0.592810	
182	2	h	0.188673	0.190930	0.999964	-1.425058	
182	2	j	-0.009834	0.009050	0.994672	-0.968134	
182	2	l	0.157129	0.158432	0.999989	-1.458326	
182	2	m	-0.198272	-0.175150	0.994150	-1.148751	
182	2	n	-0.190478	-0.159352	0.992020	-1.320885	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.143653	0.993570	-0.579173	
184	1	e	0.064221	0.095041	0.988992	-0.931832	
184	1	h	-0.045491	-0.042586	0.999765	-0.599713	
184	1	j	0.003093	0.026350	0.994689	-1.009440	
184	1	m	0.080914	0.102501	0.993606	-0.857047	
184	1	n	-0.007330	0.023365	0.992337	-1.109195	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.129955	0.986155	-0.032134	
184	2	e	0.142193	0.157319	0.994299	-0.640365	
184	2	f	0.039625	0.041512	0.986603	-0.051580	
184	2	g	0.121064	0.121775	0.985399	-0.018749	
184	2	h	-0.032087	-0.031778	0.999967	-0.169428	
184	2	j	0.038815	0.050309	0.994543	-0.492561	
184	2	l	-0.004347	-0.004208	1.000000	-1.138199	
184	2	m	0.076227	0.083962	0.990307	-0.249245	
184	2	n	-0.024803	-0.005680	0.995324	-0.884545	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.751612	0.999041	1.047436	
101	1	e	-0.338200	-0.337852	0.999810	-0.096571	
101	1	h	-0.084690	-0.079805	0.999592	-0.874708	
101	1	j	-0.605974	-0.561823	0.988776	-1.741399	
101	1	m	-0.742554	-0.746668	0.998558	0.579180	
101	1	n	-0.692001	-0.698717	0.998276	0.797335	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.700325	0.999297	-0.137691	
101	2	e	-0.339941	-0.339150	0.999428	-0.126865	
101	2	f	-0.662621	-0.662366	0.998983	-0.038515	
101	2	g	-0.701503	-0.700533	0.999215	-0.175042	
101	2	h	-0.031633	-0.031677	0.999999	0.144143	
101	2	j	-0.712239	-0.707026	0.998583	-0.698802	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.729278	0.998978	-0.388272	
101	2	n	-0.741866	-0.741931	0.998306	0.008552	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.053675	0.999990	-0.340213	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.319717	1.000000	0.549865	
103	1	j	-0.316556	-0.325296	0.997031	0.585491	
103	1	m	-0.333513	-0.331908	0.999771	-0.388994	
103	1	n	-0.327129	-0.325747	0.999723	-0.304195	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.004524	0.999986	0.739381	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.032471	0.999997	0.739374	
103	2	g	-0.008358	-0.009076	0.999988	0.716084	
103	2	h	-0.314465	-0.314490	1.000000	1.123470	
103	2	j	-0.350808	-0.356108	0.998673	0.537535	
103	2	l	-0.249678	-0.249678	1.000000	0.000000	
103	2	m	-0.213815	-0.215067	0.999865	0.381938	
103	2	n	-0.242461	-0.244068	0.999797	0.402552	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.069419	0.999888	-0.550944	
104	1	d	-0.623340	-0.624274	0.999945	0.507177	
104	1	e	-0.429965	-0.430780	0.999996	1.324326	
104	1	h	-0.347650	-0.347844	0.999996	0.324887	
104	1	j	-0.470712	-0.470223	0.999250	-0.063888	
104	1	m	-0.602159	-0.602295	0.999842	0.042853	
104	1	n	-0.415472	-0.389484	0.994046	-1.145408	
104	2	a	-0.153981	-0.152407	0.999840	-0.397725	
104	2	b	-0.161980	-0.161479	0.999965	-0.272674	
104	2	c	-0.149594	-0.148053	0.999868	-0.428003	
104	2	d	-0.592005	-0.593053	0.999968	0.722383	
104	2	e	-0.426036	-0.427124	0.999991	1.241918	
104	2	f	-0.577165	-0.577729	0.999992	0.761764	
104	2	g	-0.595353	-0.596328	0.999972	0.713880	
104	2	h	-0.178700	-0.179395	0.999995	1.020076	
104	2	j	-0.472299	-0.475769	0.999810	0.891137	
104	2	l	-0.017455	-0.017564	1.000000	1.230416	
104	2	m	-0.565612	-0.565814	0.999955	0.115405	
104	2	n	-0.384919	-0.358337	0.995359	-1.305987	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.362459	0.999991	0.775989	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.152809	0.999998	-0.657191	
107	1	j	-0.185056	-0.186681	0.999945	0.652383	
107	1	m	-0.327307	-0.329823	0.999947	1.052114	
107	1	n	-0.283932	-0.287654	0.999870	0.985711	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.357489	0.999989	0.803241	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.357901	0.999997	0.789957	
107	2	g	-0.356079	-0.356867	0.999991	0.801685	
107	2	h	0.182394	0.181926	0.999988	0.401789	
107	2	j	-0.089937	-0.092925	0.999868	0.760167	
107	2	l	0.074758	0.074394	0.999998	0.696253	
107	2	m	-0.315778	-0.317798	0.999955	0.916822	
107	2	n	-0.296199	-0.299332	0.999907	0.985588	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.361494	0.999989	1.173908	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.067282	0.999885	0.440498	
109	1	j	-0.131462	-0.115245	0.998777	-1.775086	
109	1	m	-0.328732	-0.311233	0.998681	-1.899918	
109	1	n	-0.358384	-0.330916	0.996632	-1.878653	
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.285957	0.999995	1.320403	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.294256	0.999998	1.295906	
109	2	g	-0.289566	-0.290261	0.999996	1.317980	
109	2	h	-0.068850	-0.069368	0.999962	0.319351	
109	2	j	-0.125951	-0.116716	0.999744	-2.204416	<****
109	2	l	-0.069214	-0.070142	0.998916	0.107587	
109	2	m	-0.296651	-0.281933	0.999048	-1.868668	
109	2	n	-0.330370	-0.304518	0.997145	-1.906291	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.798217	0.999967	0.792067	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.035626	0.998908	0.805472	
135	1	j	-0.633972	-0.644208	0.980286	0.276535	
135	1	m	-0.796876	-0.800037	0.999717	0.870626	
135	1	n	-0.842132	-0.811056	0.995354	-1.830233	
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.819090	0.999769	0.257540	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.818917	0.999919	-0.024199	
135	2	g	-0.816188	-0.817158	0.999824	0.366650	
135	2	h	-0.001788	-0.002120	0.999999	1.315987	
135	2	j	-0.766710	-0.778185	0.995625	0.772382	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.816586	0.999667	0.782069	
135	2	n	-0.826860	-0.797568	0.995694	-1.778553	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.257921	0.999885	0.690355	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.209714	0.780641	-0.461590	
142	1	j	-0.318605	-0.290542	0.824939	-0.218007	
142	1	m	-0.304107	-0.243396	0.954047	-0.905335	
142	1	n	-0.211922	-0.112039	0.938905	-1.263212	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.340976	0.999884	0.462161	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.453093	0.999974	0.429176	
142	2	g	-0.352275	-0.353638	0.999907	0.464213	
142	2	h	-0.277419	-0.027069	0.120421	-0.853365	
142	2	j	-0.316247	-0.335038	0.799655	0.137673	
142	2	l	-0.255325	-0.001435	0.395597	-1.038994	
142	2	m	-0.359307	-0.321676	0.959040	-0.608619	
142	2	n	-0.254817	-0.194305	0.943510	-0.804963	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.163220	0.999985	0.217269	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.268308	0.999643	1.034222	
158	1	j	-0.172847	-0.185305	0.991290	0.358875	
158	1	m	-0.212637	-0.219746	0.999828	1.451400	
158	1	n	-0.190370	-0.209590	0.992926	0.616191	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.114001	0.999868	0.007984	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.015779	0.999971	0.262813	
158	2	g	-0.089306	-0.089416	0.999898	0.028958	
158	2	h	-0.136551	-0.175037	0.997409	2.004835	(****
158	2	j	-0.103790	-0.118837	0.998761	1.136096	
158	2	l	-0.007262	-0.020390	0.999677	1.932696	
158	2	m	-0.131206	-0.137808	0.999848	1.426312	
158	2	n	-0.120778	-0.135942	0.996106	0.647920	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	d	-0.731051	-0.729330	0.999896	-0.744580	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.165300	0.999969	0.738479	
170	1	j	-0.594562	-0.595376	0.999847	0.252098	
170	1	m	-0.636259	-0.635420	0.999983	-0.789984	
170	1	n	-0.515708	-0.516671	0.999991	1.138468	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.690993	0.999972	-0.570539	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.696486	0.999991	-0.628686	
170	2	g	-0.689511	-0.688913	0.999978	-0.532944	
170	2	h	0.153761	0.153380	0.999986	0.312437	
170	2	j	-0.553947	-0.553885	0.999891	-0.022236	
170	2	l	0.045678	0.046002	0.999997	-0.622328	
170	2	m	-0.623943	-0.623468	0.999987	-0.515001	
170	2	n	-0.533728	-0.534747	0.999992	1.281938	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.204120	0.997707	-1.405551	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.185760	0.999990	-0.037143	
182	1	j	-0.008062	0.012035	0.996180	-1.216836	
182	1	m	-0.187365	-0.156366	0.992670	-1.371693	
182	1	n	-0.166005	-0.131268	0.990061	-1.316344	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.207851	0.998328	-1.446237	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.138295	0.999463	-1.457952	
182	2	g	-0.195747	-0.181985	0.998714	-1.456521	
182	2	h	0.188673	0.186046	0.999861	0.846242	
182	2	j	-0.009834	0.008408	0.995247	-0.990166	
182	2	l	0.157129	0.156170	0.999965	0.615318	
182	2	m	-0.198272	-0.172524	0.994889	-1.367066	
182	2	n	-0.190478	-0.161261	0.992680	-1.294802	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.299409	0.999723	0.416233	
180	1	d	-0.475084	-0.485908	0.999376	1.677809	
180	1	e	-0.484654	-0.484654	1.000000	0.000000	
180	1	h	-0.234297	-0.250428	0.998063	1.327498	
180	1	j	-0.395090	-0.398504	0.998871	0.390883	
180	1	m	-0.473372	-0.482114	0.999679	1.868891	
180	1	n	-0.428501	-0.432483	0.999846	1.235124	
180	2	a	-0.219548	-0.222590	0.999568	0.530523	
180	2	b	-0.157462	-0.159036	0.999904	0.573887	
180	2	c	-0.213261	-0.215989	0.999684	0.554789	
180	2	d	-0.475058	-0.487901	0.999342	1.914103	
180	2	e	-0.434632	-0.434632	1.000000	0.000000	
180	2	f	-0.393963	-0.404198	0.999606	1.916450	
180	2	g	-0.468597	-0.481172	0.999380	1.925501	
180	2	h	-0.352302	-0.377404	0.991004	0.997600	
180	2	j	-0.449142	-0.454957	0.999533	1.051901	
180	2	l	-0.126480	-0.138984	0.999041	1.437095	
180	2	m	-0.476596	-0.486384	0.999647	1.983032	(****
180	2	n	-0.445123	-0.450480	0.999766	1.353539	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.129805	0.999996	-1.072488	
184	1	e	0.064221	0.064221	1.000000	0.000000	
184	1	h	-0.045491	-0.045478	0.999958	-0.006535	
184	1	j	0.003093	0.006652	0.999801	-0.797005	
184	1	m	0.080914	0.082766	0.999986	-1.588080	
184	1	n	-0.007330	-0.002700	0.999864	-1.256942	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.129277	0.999996	-0.851096	
184	2	e	0.142193	0.142193	1.000000	0.000000	
184	2	f	0.039625	0.039735	0.999999	-0.414221	
184	2	g	0.121064	0.121514	0.999997	-0.832490	
184	2	h	-0.032087	-0.033874	0.999309	0.215014	
184	2	j	0.038815	0.039882	0.999887	-0.317027	
184	2	l	-0.004347	-0.004624	0.999980	0.195383	
184	2	m	0.076227	0.076962	0.999980	-0.516291	
184	2	n	-0.024803	-0.022960	0.999965	-0.983740	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.752322	0.997931	0.808034	
101	1	e	-0.338200	-0.341814	0.999590	0.682320	
101	1	h	-0.084690	-0.079382	0.999586	-0.943369	
101	1	j	-0.605974	-0.560459	0.986712	-1.659860	
101	1	m	-0.742554	-0.741852	0.996874	-0.067649	
101	1	n	-0.692001	-0.683252	0.996031	-0.681516	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.695983	0.997959	-0.560160	
101	2	e	-0.339941	-0.344668	0.998915	0.549529	
101	2	f	-0.662621	-0.656817	0.997499	-0.552888	
101	2	g	-0.701503	-0.695683	0.997818	-0.621423	
101	2	h	-0.031633	-0.031647	0.999998	0.040856	
101	2	j	-0.712239	-0.696749	0.997213	-1.405118	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.720449	0.997470	-1.121817	
101	2	n	-0.741866	-0.722653	0.996692	-1.620531	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.066044	0.999426	1.743908	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.319743	1.000000	1.210674	
103	1	j	-0.316556	-0.329609	0.996832	0.844967	
103	1	m	-0.333513	-0.344174	0.999021	1.241036	
103	1	n	-0.327129	-0.336017	0.999064	1.058705	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.010393	0.999842	1.837296	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.029448	0.999957	1.801669	
103	2	g	-0.008358	-0.014515	0.999866	1.844286	
103	2	h	-0.314465	-0.314500	1.000000	1.429243	
103	2	j	-0.350808	-0.358694	0.998611	0.780180	
103	2	l	-0.249678	-0.249678	1.000000	0.000000	
103	2	m	-0.213815	-0.221119	0.999677	1.433116	
103	2	n	-0.242461	-0.249939	0.999593	1.316620	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.627242	0.999709	0.905140	
104	1	e	-0.429965	-0.432473	0.999958	1.318606	
104	1	h	-0.347650	-0.347405	0.999997	-0.486152	
104	1	j	-0.470712	-0.468763	0.999504	-0.313059	
104	1	m	-0.602159	-0.602996	0.999857	0.276386	
104	1	n	-0.415472	-0.384340	0.994260	-1.385639	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.594546	0.999882	0.899312	
104	2	e	-0.426036	-0.428956	0.999934	1.235247	
104	2	f	-0.577165	-0.578412	0.999972	0.888252	
104	2	g	-0.595353	-0.597809	0.999891	0.908231	
104	2	h	-0.178700	-0.179124	0.999996	0.679858	
104	2	j	-0.472299	-0.476056	0.999869	1.151360	
104	2	l	-0.017455	-0.017455	1.000000	0.000000	
104	2	m	-0.565612	-0.566576	0.999931	0.443463	
104	2	n	-0.384919	-0.357451	0.995310	-1.341035	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.362459	0.999991	0.775989	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.152809	0.999998	-0.657191	
107	1	j	-0.185056	-0.186681	0.999945	0.652383	
107	1	m	-0.327307	-0.329823	0.999947	1.052114	
107	1	n	-0.283932	-0.287654	0.999870	0.985711	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.357489	0.999989	0.803241	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.357901	0.999997	0.789957	
107	2	g	-0.356079	-0.356867	0.999991	0.801685	
107	2	h	0.182394	0.181926	0.999988	0.401789	
107	2	j	-0.089937	-0.092925	0.999868	0.760167	
107	2	l	0.074758	0.074394	0.999998	0.696253	
107	2	m	-0.315778	-0.317798	0.999955	0.916822	
107	2	n	-0.296199	-0.299332	0.999907	0.985588	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.219931	0.999868	-0.199317	
109	1	d	-0.360544	-0.360360	0.999729	-0.045736	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.066351	0.999942	0.152409	
109	1	j	-0.131462	-0.122050	0.999671	-1.984724	(****
109	1	m	-0.328732	-0.323165	0.997027	-0.410927	
109	1	n	-0.358384	-0.342772	0.995164	-0.908314	
109	2	a	-0.228134	-0.227383	0.999846	-0.237046	
109	2	b	-0.233510	-0.233074	0.999920	-0.190643	
109	2	c	-0.234593	-0.233846	0.999866	-0.253052	
109	2	d	-0.285177	-0.286415	0.999469	0.213575	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.294547	0.999285	0.105794	
109	2	g	-0.289566	-0.290706	0.999402	0.185584	
109	2	h	-0.068850	-0.068648	0.999993	-0.284085	
109	2	j	-0.125951	-0.120357	0.999848	-1.736169	
109	2	l	-0.069214	-0.069423	0.999747	0.050043	
109	2	m	-0.296651	-0.281066	0.998933	-1.869049	
109	2	n	-0.330370	-0.301879	0.996543	-1.908738	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.772840	0.991218	-1.140052	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.035609	0.998909	0.803948	
135	1	j	-0.633972	-0.637440	0.980311	0.093623	
135	1	m	-0.796876	-0.778377	0.996115	-1.275569	
135	1	n	-0.842132	-0.795701	0.992775	-2.013543	(****
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.809892	0.995072	-0.591501	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.817350	0.998138	-0.188927	
135	2	g	-0.816188	-0.807974	0.996205	-0.650900	
135	2	h	-0.001788	-0.002120	0.999999	1.315987	
135	2	j	-0.766710	-0.761965	0.993959	-0.275507	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.807226	0.996486	-0.530789	
135	2	n	-0.826860	-0.797037	0.995023	-1.717193	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.269432	0.999431	1.819247	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.209714	0.780641	-0.461590	
142	1	j	-0.318605	-0.291969	0.824910	-0.206963	
142	1	m	-0.304107	-0.254739	0.955165	-0.747293	
142	1	n	-0.211922	-0.120100	0.940354	-1.175800	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.346993	0.999751	1.532558	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.455626	0.999944	1.419687	
142	2	g	-0.352275	-0.359016	0.999799	1.531634	
142	2	h	-0.277419	-0.027069	0.120421	-0.853365	
142	2	j	-0.316247	-0.338219	0.799887	0.161175	
142	2	l	-0.255325	-0.001435	0.395597	-1.038994	
142	2	m	-0.359307	-0.327471	0.959547	-0.519066	
142	2	n	-0.254817	-0.198757	0.944209	-0.750871	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.168945	0.999338	0.628346	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.268266	0.999657	1.047834	
158	1	j	-0.172847	-0.188528	0.991179	0.448889	
158	1	m	-0.212637	-0.224449	0.999513	1.436359	
158	1	n	-0.190370	-0.210847	0.992773	0.649489	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.127840	0.996101	0.591873	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.000194	0.996324	0.703471	
158	2	g	-0.089306	-0.104073	0.995964	0.617744	
158	2	h	-0.136551	-0.174185	0.997516	2.002268	(****
158	2	j	-0.103790	-0.126435	0.997999	1.345056	
158	2	l	-0.007262	-0.020129	0.999681	1.904769	
158	2	m	-0.131206	-0.148421	0.997938	1.010909	
158	2	n	-0.120778	-0.138368	0.995899	0.732335	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	c	-0.731051	-0.729927	0.999868	-0.437500	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.165325	0.999969	0.718610	
170	1	j	-0.594562	-0.600554	0.998838	0.667406	
170	1	m	-0.636259	-0.636255	0.999905	-0.001826	
170	1	n	-0.515708	-0.518841	0.999802	0.790715	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.691445	0.999960	-0.177914	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.696701	0.999988	-0.288599	
170	2	g	-0.689511	-0.689362	0.999967	-0.110846	
170	2	h	0.153761	0.153339	0.999985	0.343563	
170	2	j	-0.553947	-0.555406	0.999790	0.372142	
170	2	i	0.045678	0.045931	0.999997	-0.476422	
170	2	m	-0.623943	-0.624053	0.999959	0.067787	
170	2	n	-0.533728	-0.535895	0.999932	0.941700	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.299409	0.999723	0.416233	
180	1	d	-0.475084	-0.485505	0.996877	0.746074	
180	1	e	-0.484654	-0.483529	0.999452	-0.194078	
180	1	h	-0.234297	-0.250120	0.998018	1.287587	
180	1	j	-0.395090	-0.397840	0.997360	0.206057	
180	1	m	-0.473372	-0.482802	0.996953	0.683441	
180	1	n	-0.428501	-0.436343	0.997985	0.681462	
180	2	a	-0.219548	-0.222590	0.999568	0.530523	
180	2	b	-0.157462	-0.159036	0.999904	0.573887	
180	2	c	-0.213261	-0.215989	0.999684	0.554789	
180	2	d	-0.475058	-0.497852	0.998381	2.139282	<****
180	2	e	-0.434632	-0.431416	0.998580	-0.334327	
180	2	f	-0.393963	-0.417118	0.998748	2.385579	<****
180	2	g	-0.468597	-0.492359	0.998440	2.249597	<****
180	2	h	-0.352302	-0.377037	0.990957	0.980613	
180	2	j	-0.449142	-0.459918	0.999212	1.482148	
180	2	l	-0.126480	-0.138760	0.999033	1.405832	
180	2	m	-0.476596	-0.494443	0.998821	1.982610	<****
180	2	n	-0.445123	-0.457971	0.998923	1.508164	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.220375	0.999972	-1.021391	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.184857	0.999994	1.386589	
182	1	j	-0.008062	-0.013225	0.999832	1.491479	
182	1	m	-0.187365	-0.186776	0.999827	-0.170309	
182	1	n	-0.166005	-0.167819	0.999863	0.587933	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.221788	0.999982	-1.432104	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.146781	0.999998	-1.400394	
182	2	g	-0.195747	-0.194583	0.999990	-1.385283	
182	2	h	0.188673	0.184829	0.999864	1.251626	
182	2	j	-0.009834	-0.018578	0.999581	1.598529	
182	2	l	0.157129	0.155412	0.999968	1.140827	
182	2	m	-0.198272	-0.197672	0.999900	-0.228425	
182	2	n	-0.190478	-0.192268	0.999884	0.633289	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.127115	0.999601	0.316625	
184	1	e	0.064221	0.062900	0.999970	0.768734	
184	1	h	-0.045491	-0.045629	0.999958	0.067423	
184	1	j	0.003093	0.006609	0.998948	-0.342742	
184	1	m	0.080914	0.080696	0.998909	0.020933	
184	1	n	-0.007330	-0.003656	0.999516	-0.527994	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.128098	0.999892	0.206697	
184	2	e	0.142193	0.140976	0.999974	0.763283	
184	2	f	0.039625	0.039648	0.999976	-0.014669	
184	2	g	0.121064	0.120505	0.999914	0.192686	
184	2	h	-0.032087	-0.034065	0.999303	0.237118	
184	2	j	0.038815	0.039566	0.999590	-0.117423	
184	2	l	-0.004347	-0.004624	0.999980	0.195383	
184	2	m	0.076227	0.076297	0.999811	-0.016219	
184	2	n	-0.024803	-0.023155	0.999903	-0.527956	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.751612	0.999041	1.047436	
101	1	e	-0.338200	-0.337852	0.999810	-0.096571	
101	1	h	-0.084690	-0.079805	0.999592	-0.874708	
101	1	j	-0.605974	-0.561823	0.988776	-1.741399	
101	1	m	-0.742554	-0.746668	0.998558	0.579180	
101	1	n	-0.692001	-0.698717	0.998276	0.797335	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.700325	0.999297	-0.137691	
101	2	e	-0.339941	-0.339150	0.999428	-0.126865	
101	2	f	-0.662621	-0.662366	0.998983	-0.038515	
101	2	g	-0.701503	-0.700533	0.999215	-0.175042	
101	2	h	-0.031633	-0.031677	0.999999	0.144143	
101	2	j	-0.712239	-0.707026	0.998583	-0.698802	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.729278	0.998978	-0.388272	
101	2	n	-0.741866	-0.741931	0.998306	0.008552	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.051487	0.999946	-1.182141	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.320319	0.999982	0.527310	
103	1	j	-0.316556	-0.326918	0.996357	0.626573	
103	1	m	-0.333513	-0.327510	0.999200	-0.817553	
103	1	n	-0.327129	-0.321477	0.999546	-0.965824	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.003883	0.999982	0.126814	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.032771	0.999996	0.142771	
103	2	g	-0.008358	-0.008464	0.999984	0.091623	
103	2	h	-0.314465	-0.315726	0.999959	0.717689	
103	2	j	-0.350808	-0.356913	0.998329	0.551869	
103	2	l	-0.249678	-0.249491	0.999991	-0.220407	
103	2	m	-0.213815	-0.213207	0.999812	-0.157101	
103	2	n	-0.242461	-0.242560	0.999784	0.024196	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.623721	0.999985	0.401482	
104	1	e	-0.429965	-0.430780	0.999996	1.324326	
104	1	h	-0.347650	-0.347405	0.999997	-0.486152	
104	1	j	-0.470712	-0.467840	0.999363	-0.406482	
104	1	m	-0.602159	-0.600803	0.999930	-0.636572	
104	1	n	-0.415472	-0.383766	0.994267	-1.410728	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.592818	0.999973	0.605450	
104	2	e	-0.426036	-0.427124	0.999991	1.241918	
104	2	f	-0.577165	-0.577596	0.999993	0.633143	
104	2	g	-0.595353	-0.596113	0.999976	0.600343	
104	2	h	-0.178700	-0.179124	0.999996	0.679858	
104	2	j	-0.472299	-0.475148	0.999825	0.766603	
104	2	l	-0.017455	-0.017455	1.000000	0.000000	
104	2	m	-0.565612	-0.565380	0.999964	-0.148670	
104	2	n	-0.384919	-0.356937	0.995318	-1.366158	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.366005	0.999845	1.064491	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.152809	0.999998	-0.657191	
107	1	j	-0.185056	-0.187224	0.999938	0.815546	
107	1	m	-0.327307	-0.331679	0.999883	1.231026	
107	1	n	-0.283932	-0.288712	0.999849	1.171721	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.358899	0.999966	1.199868	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.358801	0.999988	1.194389	
107	2	g	-0.356079	-0.358155	0.999972	1.198332	
107	2	h	0.182394	0.181926	0.999988	0.401789	
107	2	j	-0.089937	-0.093350	0.999861	0.846109	
107	2	i	0.074758	0.074394	0.999998	0.696253	
107	2	m	-0.315778	-0.318551	0.999941	1.099238	
107	2	n	-0.296199	-0.299772	0.999901	1.084805	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	p > .05
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.360341	0.999978	-0.176233	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.066833	0.999878	0.271868	
109	1	j	-0.131462	-0.115245	0.998777	-1.775086	
109	1	m	-0.328732	-0.307707	0.998562	-2.172109	<****
109	1	n	-0.358384	-0.327905	0.996556	-2.051845	<****
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.285164	0.999990	-0.016233	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.293866	0.999997	0.069221	
109	2	g	-0.289566	-0.289553	0.999992	-0.017410	
109	2	h	-0.068850	-0.069122	0.999959	0.162325	
109	2	j	-0.125951	-0.116716	0.999744	-2.204416	<****
109	2	l	-0.069214	-0.069945	0.998914	0.084582	
109	2	m	-0.296651	-0.280523	0.999037	-2.030152	<****
109	2	n	-0.330370	-0.303095	0.997139	-2.004619	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.798253	0.999966	0.806227	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.035626	0.998908	0.805472	
135	1	j	-0.633972	-0.644212	0.980281	0.276614	
135	1	m	-0.796876	-0.800067	0.999716	0.876231	
135	1	n	-0.842132	-0.811076	0.995357	-1.829803	
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.819110	0.999765	0.261981	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.818926	0.999918	-0.019273	
135	2	g	-0.816188	-0.817176	0.999821	0.370452	
135	2	h	-0.001788	-0.002120	0.999999	1.315987	
135	2	j	-0.766710	-0.778213	0.995618	0.773603	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.816604	0.999662	0.781532	
135	2	n	-0.826860	-0.797563	0.995690	-1.778178	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.257921	0.999885	0.690355	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.209714	0.780641	-0.461590	
142	1	j	-0.318605	-0.290542	0.824939	-0.218007	
142	1	m	-0.304107	-0.243396	0.954047	-0.905335	
142	1	n	-0.211922	-0.112039	0.938905	-1.263212	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.340976	0.999884	0.462161	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.453093	0.999974	0.429176	
142	2	g	-0.352275	-0.353638	0.999907	0.464213	
142	2	h	-0.277419	-0.027069	0.120421	-0.853365	
142	2	j	-0.316247	-0.335038	0.799655	0.137673	
142	2	l	-0.255325	-0.001435	0.395597	-1.038994	
142	2	m	-0.359307	-0.321676	0.959040	-0.608619	
142	2	n	-0.254817	-0.194305	0.943510	-0.804963	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	c	-0.162911	-0.164549	0.999945	0.593286	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.268308	0.999643	1.034222	
158	1	j	-0.172847	-0.187373	0.991239	0.417235	
158	1	m	-0.212637	-0.221337	0.999757	1.496819	
158	1	n	-0.190370	-0.210652	0.992794	0.644254	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.113303	0.999831	-0.136053	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.016700	0.999829	-0.078035	
158	2	g	-0.089306	-0.088465	0.999836	-0.174200	
158	2	h	-0.136551	-0.175042	0.997408	2.004895	(****
158	2	j	-0.103790	-0.119243	0.998767	1.169869	
158	2	l	-0.007262	-0.020390	0.999677	1.932696	
158	2	m	-0.131206	-0.137477	0.999837	1.307669	
158	2	n	-0.120778	-0.136421	0.996092	0.667111	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	d	-0.731051	-0.742090	0.993649	0.621921	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.164412	0.999960	1.089657	
170	1	j	-0.594562	-0.614372	0.990243	0.764024	
170	1	m	-0.636259	-0.654478	0.992729	0.842953	
170	1	n	-0.515708	-0.541451	0.991535	0.994660	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.707218	0.993231	0.794390	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.713767	0.989529	0.703046	
170	2	g	-0.689511	-0.705860	0.992551	0.796411	
170	2	h	0.153761	0.152558	0.999978	0.801444	
170	2	j	-0.553947	-0.575755	0.988949	0.764538	
170	2	l	0.045678	0.045717	0.999997	-0.067077	
170	2	m	-0.623943	-0.644082	0.991707	0.861803	
170	2	n	-0.533728	-0.560072	0.990025	0.951460	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.300069	0.999718	0.558162	
180	1	d	-0.475084	-0.469676	0.997561	-0.438159	
180	1	e	-0.484654	-0.484654	1.000000	0.000000	
180	1	h	-0.234297	-0.250499	0.998062	1.332840	
180	1	j	-0.395090	-0.391709	0.998332	-0.318095	
180	1	m	-0.473372	-0.469709	0.998407	-0.367378	
180	1	n	-0.428501	-0.426252	0.999451	-0.374795	
180	2	a	-0.219548	-0.223547	0.999558	0.689100	
180	2	b	-0.157462	-0.159451	0.999902	0.719927	
180	2	c	-0.213261	-0.216810	0.999676	0.713390	
180	2	d	-0.475058	-0.463184	0.996392	-0.785163	
180	2	e	-0.434632	-0.434632	1.000000	0.000000	
180	2	f	-0.393963	-0.372357	0.995482	-1.214745	
180	2	g	-0.468597	-0.454845	0.996089	-0.868267	
180	2	h	-0.352302	-0.377455	0.991003	0.999521	
180	2	j	-0.449142	-0.447423	0.999231	-0.245109	
180	2	l	-0.126480	-0.138939	0.999040	1.431369	
180	2	m	-0.476596	-0.473019	0.998716	-0.400072	
180	2	n	-0.445123	-0.445674	0.999652	0.116776	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.204120	0.997707	-1.405551	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.185760	0.999990	-0.037143	
182	1	j	-0.008062	0.012035	0.996180	-1.216836	
182	1	m	-0.187365	-0.156366	0.992670	-1.371693	
182	1	n	-0.166005	-0.131268	0.990061	-1.316344	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.207851	0.998328	-1.446237	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.138295	0.999463	-1.457952	
182	2	g	-0.195747	-0.181985	0.998714	-1.456521	
182	2	h	0.188673	0.186046	0.999861	0.846242	
182	2	j	-0.009834	0.008408	0.995247	-0.990166	
182	2	i	0.157129	0.156170	0.999965	0.615318	
182	2	m	-0.198272	-0.172524	0.994889	-1.367066	
182	2	n	-0.190478	-0.161261	0.992680	-1.294802	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.136537	0.999522	-1.083389	
184	1	e	0.064221	0.064221	1.000000	0.000000	
184	1	h	-0.045491	-0.045478	0.999958	-0.006535	
184	1	j	0.003093	0.008376	0.999761	-1.079631	
184	1	m	0.080914	0.087276	0.999714	-1.193218	
184	1	n	-0.007330	-0.002087	0.999859	-1.396917	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.133600	0.999816	-1.133947	
184	2	e	0.142193	0.142193	1.000000	0.000000	
184	2	f	0.039625	0.042174	0.999919	-0.897806	
184	2	g	0.121064	0.125731	0.999819	-1.103698	
184	2	h	-0.032087	-0.033874	0.999309	0.215014	
184	2	j	0.038815	0.041741	0.999831	-0.712119	
184	2	l	-0.004347	-0.004624	0.999980	0.195383	
184	2	m	0.076227	0.079961	0.999855	-0.982313	
184	2	n	-0.024803	-0.022576	0.999962	-1.138266	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.750839	0.999025	0.916500	
101	1	e	-0.338200	-0.420617	0.951057	1.433873	
101	1	h	-0.084690	-0.072247	0.999203	-1.593022	
101	1	j	-0.605974	-0.553144	0.985667	-1.827722	
101	1	m	-0.742554	-0.740466	0.997804	-0.239187	
101	1	n	-0.692001	-0.677530	0.991948	-0.785997	
101	2	a	-0.381796	-0.379436	0.999941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.710868	0.998401	1.197970	
101	2	e	-0.339941	-0.413889	0.960234	1.424931	
101	2	f	-0.662621	-0.675150	0.997504	1.173007	
101	2	g	-0.701503	-0.711851	0.998177	1.184736	
101	2	h	-0.031633	-0.031561	0.999999	-0.226240	
101	2	j	-0.712239	-0.711003	0.998551	-0.166544	
101	2	l	-0.001532	-0.001532	1.000000	0.000000	
101	2	m	-0.731639	-0.736005	0.998631	0.619011	
101	2	n	-0.741866	-0.736241	0.997741	-0.626018	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
103	1	a	0.146284	0.144978	0.999953	0.665002	
103	1	d	-0.053992	-0.080883	0.994353	1.242260	
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.319721	1.000000	0.111545	
103	1	j	-0.316556	-0.332959	0.996318	0.984024	
103	1	m	-0.333513	-0.353968	0.992094	0.844509	
103	1	n	-0.327129	-0.342995	0.991868	0.645213	
103	2	a	0.257744	0.256450	0.999941	0.602435	
103	2	b	0.295230	0.294689	0.999988	0.563265	
103	2	c	0.256569	0.255414	0.999951	0.592566	
103	2	d	-0.003726	-0.022769	0.997177	1.241748	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.016124	0.997474	1.153621	
103	2	g	-0.008358	-0.028092	0.996915	1.231121	
103	2	h	-0.314465	-0.313860	0.999993	-0.822045	
103	2	j	-0.350808	-0.363150	0.997681	0.943716	
103	2	l	-0.249678	-0.249466	1.000000	-1.329759	
103	2	m	-0.213815	-0.235891	0.994566	1.061138	
103	2	n	-0.242461	-0.263263	0.993529	0.923036	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.622474	0.999659	-0.189493	
104	1	e	-0.429965	-0.431091	0.999992	1.323283	
104	1	h	-0.347650	-0.368062	0.996505	1.154530	
104	1	j	-0.470712	-0.476799	0.997358	0.424038	
104	1	m	-0.602159	-0.605039	0.998952	0.351807	
104	1	n	-0.415472	-0.396591	0.988208	-0.599432	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.604637	0.997679	1.006654	
104	2	e	-0.426036	-0.427762	0.999977	1.239605	
104	2	f	-0.577165	-0.594427	0.996423	1.091747	
104	2	g	-0.595353	-0.608863	0.997278	0.997852	
104	2	h	-0.178700	-0.184904	0.999760	1.281626	
104	2	j	-0.472299	-0.492604	0.997362	1.379610	
104	2	l	-0.017455	-0.017892	0.999998	1.079241	
104	2	m	-0.565612	-0.581620	0.997540	1.201384	
104	2	n	-0.384919	-0.377151	0.988233	-0.244910	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.374103	0.992962	0.462016	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.132720	0.994950	0.820388	
107	1	j	-0.185056	-0.223991	0.989567	1.130729	
107	1	m	-0.327307	-0.343330	0.991538	0.537701	
107	1	n	-0.283932	-0.302328	0.994482	0.752152	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.354579	0.997639	-0.130842	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.353567	0.998151	-0.283723	
107	2	g	-0.356079	-0.353170	0.997673	-0.187974	
107	2	h	0.182394	0.172443	0.997623	0.604073	
107	2	j	-0.089937	-0.110899	0.997084	1.136444	
107	2	l	0.074758	0.067997	0.999509	0.891345	
107	2	m	-0.315778	-0.318261	0.998629	0.206098	
107	2	n	-0.296199	-0.304222	0.998930	0.746373	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.397370	0.987902	1.360074	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.067439	0.999885	0.496746	
109	1	j	-0.131462	-0.120445	0.998601	-1.129571	
109	1	m	-0.328732	-0.334144	0.994099	0.284255	
109	1	n	-0.358384	-0.339139	0.995641	-1.174331	
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.323096	0.985664	1.257465	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.328449	0.988789	1.299676	
109	2	g	-0.289566	-0.327214	0.985987	1.264255	
109	2	n	-0.068850	-0.069508	0.999962	0.404447	
109	2	j	-0.125951	-0.121717	0.999514	-0.736985	
109	2	l	-0.069214	-0.070241	0.998916	0.119081	
109	2	m	-0.296651	-0.306993	0.992316	0.470836	
109	2	n	-0.330370	-0.316870	0.993654	-0.680503	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600904	0.999613	1.704237	
135	1	d	-0.797236	-0.798217	0.999967	0.792067	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.035626	0.998908	0.805472	
135	1	j	-0.633972	-0.644208	0.980286	0.276535	
135	1	m	-0.796876	-0.800037	0.999717	0.870626	
135	1	n	-0.842132	-0.811056	0.995354	-1.830233	
135	2	a	-0.637630	-0.648359	0.999258	1.390371	
135	2	b	-0.647093	-0.652380	0.999766	1.247222	
135	2	c	-0.635240	-0.644584	0.999466	1.418114	
135	2	d	-0.818316	-0.819090	0.999769	0.257540	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.818917	0.999319	-0.024199	
135	2	g	-0.816188	-0.817158	0.999824	0.366650	
135	2	h	-0.001788	-0.002120	0.999999	1.315987	
135	2	j	-0.766710	-0.778185	0.995625	0.772382	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.816586	0.999667	0.782069	
135	2	n	-0.826860	-0.797568	0.995694	-1.778553	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.266719	0.998654	0.963072	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.209742	0.780646	-0.461410	
142	1	j	-0.318605	-0.299714	0.819027	-0.144612	
142	1	m	-0.304107	-0.254072	0.949614	-0.714693	
142	1	n	-0.211922	-0.123314	0.936468	-1.099862	
142	2	a	-0.324310	-0.321402	0.999345	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.339884	0.999135	0.047741	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.450467	0.998652	-0.186777	
142	2	g	-0.352275	-0.351640	0.998953	-0.064632	
142	2	h	-0.277419	-0.027069	0.120421	-0.853365	
142	2	j	-0.316247	-0.338458	0.797386	0.161938	
142	2	l	-0.255325	-0.001435	0.395597	-1.038994	
142	2	m	-0.359307	-0.322414	0.957514	-0.586120	
142	2	n	-0.254817	-0.198763	0.942636	-0.740485	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.167314	0.999524	0.540739	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.268266	0.999657	1.047817	
158	1	j	-0.172847	-0.186749	0.991273	0.400075	
158	1	m	-0.212637	-0.222111	0.999715	1.503621	
158	1	n	-0.190370	-0.209259	0.992913	0.605052	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.125940	0.996291	0.523739	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.000695	0.996384	0.687294	
158	2	g	-0.089306	-0.102465	0.996139	0.562771	
158	2	h	-0.136551	-0.174153	0.997518	2.001510	(****
158	2	j	-0.103790	-0.124319	0.998090	1.248288	
158	2	l	-0.007262	-0.020129	0.999681	1.904769	
158	2	m	-0.131206	-0.146276	0.998073	0.915683	
158	2	n	-0.120778	-0.136574	0.996080	0.672644	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	d	-0.731051	-0.731320	0.999647	0.064753	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.165329	0.999969	0.722759	
170	1	j	-0.594562	-0.595450	0.998884	0.101955	
170	1	m	-0.636259	-0.635176	0.999509	-0.195144	
170	1	n	-0.515708	-0.512295	0.999210	-0.434523	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.693355	0.999701	0.404417	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.697367	0.999305	0.188628	
170	2	g	-0.689511	-0.691282	0.999758	0.480403	
170	2	h	0.153761	0.153635	0.999985	0.102754	
170	2	j	-0.553947	-0.552176	0.999329	-0.252509	
170	2	l	0.045678	0.046002	0.999997	-0.622328	
170	2	m	-0.623943	-0.624003	0.999621	0.011997	
170	2	n	-0.533728	-0.532179	0.999460	-0.242681	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.299409	0.999723	0.416233	
180	1	d	-0.475084	-0.485908	0.999376	1.677809	
180	1	e	-0.484654	-0.484654	1.000000	0.000000	
180	1	h	-0.234297	-0.250428	0.998063	1.327498	
180	1	j	-0.395090	-0.398504	0.998871	0.390883	
180	1	m	-0.473372	-0.482114	0.999679	1.868891	
180	1	n	-0.428501	-0.432483	0.999846	1.235124	
180	2	a	-0.219548	-0.222590	0.999568	0.530523	
180	2	b	-0.157462	-0.159036	0.999904	0.573887	
180	2	c	-0.213261	-0.215989	0.999684	0.554789	
180	2	d	-0.475058	-0.487901	0.999342	1.914103	
180	2	e	-0.434632	-0.434632	1.000000	0.000000	
180	2	f	-0.393963	-0.404198	0.999606	1.916450	
180	2	g	-0.468597	-0.481172	0.999380	1.925501	
180	2	h	-0.352302	-0.377404	0.991004	0.997600	
180	2	j	-0.449142	-0.454957	0.999533	1.051901	
180	2	l	-0.126480	-0.138984	0.999041	1.437095	
180	2	m	-0.476596	-0.486384	0.999647	1.983032	<****
180	2	n	-0.445123	-0.450480	0.999766	1.353539	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.224193	0.996365	0.153693	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.185081	0.999996	1.301022	
182	1	j	-0.008062	-0.005805	0.998273	-0.203242	
182	1	m	-0.187365	-0.184546	0.996166	-0.173346	
182	1	n	-0.166005	-0.166631	0.997851	0.051283	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.221232	0.994130	-0.106675	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.147663	0.998098	0.035948	
182	2	g	-0.195747	-0.194440	0.995430	-0.073752	
182	2	h	0.188673	0.185455	0.999907	1.267652	
182	2	j	-0.009834	-0.016492	0.997645	0.513395	
182	2	l	0.157129	0.155712	0.999978	1.131076	
182	2	m	-0.198272	-0.196727	0.995693	-0.089856	
182	2	n	-0.190478	-0.193120	0.998452	0.256058	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.129016	0.999984	0.065318	
184	1	e	0.064221	0.063340	0.999987	0.768379	
184	1	h	-0.045491	-0.045483	0.999958	-0.003716	
184	1	j	0.003093	0.006338	0.999797	-0.719477	
184	1	m	0.080914	0.081955	0.999973	-0.636850	
184	1	n	-0.007330	-0.003280	0.999852	-1.053561	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.128531	0.999986	0.206309	
184	2	e	0.142193	0.140136	0.999926	0.764816	
184	2	f	0.039625	0.039450	0.999998	0.373604	
184	2	g	0.121064	0.120839	0.999989	0.215494	
184	2	h	-0.032087	-0.033915	0.999308	0.219789	
184	2	j	0.038815	0.039314	0.999870	-0.138526	
184	2	l	-0.004347	-0.004624	0.999980	0.195383	
184	2	m	0.076227	0.076291	0.999966	-0.035013	
184	2	n	-0.024803	-0.023447	0.999956	-0.645791	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
101	1	a	-0.424357	-0.422512	0.999964	-1.198307	
101	1	d	-0.745425	-0.735126	0.978804	-0.380127	
101	1	e	-0.338200	-0.354227	0.997701	1.270456	
101	1	h	-0.084690	-0.078579	0.998527	-0.575966	
101	1	j	-0.605974	-0.526366	0.947036	-1.465905	
101	1	m	-0.742554	-0.717269	0.972555	-0.795915	
101	1	n	-0.692001	-0.659707	0.968759	-0.881743	
101	2	a	-0.381796	-0.379436	0.995941	-1.188190	
101	2	b	-0.328840	-0.327643	0.999988	-1.289199	
101	2	c	-0.381283	-0.379214	0.999955	-1.185281	
101	2	d	-0.701048	-0.682226	0.981310	-0.681521	
101	2	e	-0.339941	-0.363183	0.995688	1.345865	
101	2	f	-0.662621	-0.658407	0.974987	-0.128532	
101	2	g	-0.701503	-0.683738	0.980016	-0.624398	
101	2	h	-0.031633	-0.030426	0.999290	-0.163398	
101	2	j	-0.712239	-0.660565	0.978850	-1.638745	
101	2	l	-0.001532	-0.001474	0.999999	-0.185714	
101	2	m	-0.731639	-0.696171	0.981391	-1.280029	
101	2	n	-0.741866	-0.686338	0.983399	-1.964468	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	p > .05
103	1	a	0.146284	0.151090	0.999577	-0.817224	
103	1	d	-0.053992	0.061290	0.960561	-2.020354	<****
103	1	e	-0.000528	-0.000528	1.000000	0.000000	
103	1	h	-0.319705	-0.331746	0.987033	0.387229	
103	1	j	-0.316556	-0.265306	0.961034	-0.937486	
103	1	m	-0.333513	-0.194244	0.941728	-2.050990	<****
103	1	n	-0.327129	-0.207268	0.959997	-2.123687	<****
103	2	a	0.257744	0.261709	0.999437	-0.598744	
103	2	b	0.295230	0.297466	0.999797	-0.567829	
103	2	c	0.256569	0.260124	0.999546	-0.597584	
103	2	d	-0.003726	0.079665	0.973030	-1.764198	
103	2	e	-0.000528	-0.000528	1.000000	0.000000	
103	2	f	0.032856	0.083885	0.979525	-1.238270	
103	2	g	-0.008358	0.072788	0.971869	-1.680514	
103	2	h	-0.314465	-0.327134	0.994500	0.623337	
103	2	j	-0.350808	-0.309699	0.980181	-1.064740	
103	2	i	-0.249678	-0.247160	0.998752	-0.254841	
103	2	m	-0.213815	-0.119610	0.965732	-1.784229	
103	2	n	-0.242461	-0.158955	0.972920	-1.785058	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
104	1	a	-0.071262	-0.071204	0.999904	-0.018778	
104	1	d	-0.623340	-0.547688	0.940749	-1.181029	
104	1	e	-0.429965	-0.402867	0.980538	-0.672547	
104	1	h	-0.347650	-0.383169	0.987289	1.059273	
104	1	j	-0.470712	-0.324632	0.943542	-2.035548	(****
104	1	m	-0.602159	-0.523198	0.955703	-1.380366	
104	1	n	-0.415472	-0.325862	0.966071	-1.618661	
104	2	a	-0.153981	-0.154278	0.999859	0.080141	
104	2	b	-0.161980	-0.162270	0.999969	0.165467	
104	2	c	-0.149594	-0.149793	0.999884	0.059193	
104	2	d	-0.592005	-0.479979	0.948125	-1.738371	
104	2	e	-0.426036	-0.372460	0.960155	-0.919225	
104	2	f	-0.577165	-0.439319	0.938613	-1.919253	
104	2	g	-0.595353	-0.475510	0.943964	-1.784031	
104	2	h	-0.178700	-0.217979	0.990862	1.319833	
104	2	j	-0.472299	-0.317582	0.954676	-2.360547	(****
104	2	l	-0.017455	-0.023259	0.999859	1.545265	
104	2	m	-0.565612	-0.443978	0.956310	-1.985878	(****
104	2	n	-0.384919	-0.261436	0.968198	-2.236528	(****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
107	1	a	-0.269170	-0.269170	1.000000	0.000000	
107	1	d	-0.361714	-0.435007	0.964499	1.215473	
107	1	e	-0.001095	-0.001095	1.000000	0.000000	
107	1	h	0.152533	0.168742	0.933024	-0.185090	
107	1	j	-0.185056	-0.254454	0.975135	1.308438	
107	1	m	-0.327307	-0.424017	0.929626	1.136070	
107	1	n	-0.283932	-0.416889	0.902309	1.313209	
107	2	a	-0.290729	-0.290729	1.000000	0.000000	
107	2	b	-0.285007	-0.285007	1.000000	0.000000	
107	2	c	-0.292260	-0.292260	1.000000	0.000000	
107	2	d	-0.356617	-0.448451	0.969943	1.629910	
107	2	e	-0.001095	-0.001095	1.000000	0.000000	
107	2	f	-0.357481	-0.450159	0.975357	1.800262	
107	2	g	-0.356079	-0.450427	0.970253	1.679596	
107	2	h	0.182394	0.209065	0.952498	-0.363918	
107	2	j	-0.089937	-0.184122	0.984029	2.183024	<****
107	2	l	0.074758	0.075260	0.995797	-0.022643	
107	2	m	-0.315778	-0.430255	0.961571	1.774877	
107	2	n	-0.296199	-0.411616	0.968664	1.957071	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
109	1	a	-0.220519	-0.220519	1.000000	0.000000	
109	1	d	-0.360544	-0.394514	0.987746	1.248560	
109	1	e	-0.000545	-0.000545	1.000000	0.000000	
109	1	h	-0.066046	-0.066349	0.999959	0.181646	
109	1	j	-0.131462	-0.118022	0.996936	-0.931230	
109	1	m	-0.328732	-0.349701	0.988668	0.794758	
109	1	n	-0.358384	-0.377749	0.984920	0.644409	
109	2	a	-0.228134	-0.228134	1.000000	0.000000	
109	2	b	-0.233510	-0.233510	1.000000	0.000000	
109	2	c	-0.234593	-0.234593	1.000000	0.000000	
109	2	d	-0.285177	-0.305835	0.987646	0.739346	
109	2	e	-0.000545	-0.000545	1.000000	0.000000	
109	2	f	-0.293837	-0.305177	0.988468	0.421241	
109	2	g	-0.289566	-0.307594	0.987425	0.640507	
109	2	h	-0.068850	-0.068934	0.999958	0.049112	
109	2	j	-0.125951	-0.126579	0.998624	0.064993	
109	2	l	-0.069214	-0.070141	0.998880	0.105703	
109	2	m	-0.296651	-0.307070	0.987574	0.373192	
109	2	n	-0.330370	-0.336207	0.984358	0.188531	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
135	1	a	-0.590694	-0.600528	0.999565	1.573369	
135	1	d	-0.797236	-0.802837	0.992671	0.317165	
135	1	e	-0.001234	-0.001234	1.000000	0.000000	
135	1	h	-0.026503	-0.054399	0.997769	1.723191	
135	1	j	-0.633972	-0.667409	0.958150	0.623624	
135	1	m	-0.796876	-0.807115	0.989432	0.482032	
135	1	n	-0.842132	-0.819953	0.987478	-0.983305	
135	2	a	-0.637630	-0.648033	0.999177	1.293648	
135	2	b	-0.647093	-0.652253	0.999750	1.184265	
135	2	c	-0.635240	-0.644278	0.999399	1.309125	
135	2	d	-0.818316	-0.832591	0.991892	0.791085	
135	2	e	-0.001234	-0.001234	1.000000	0.000000	
135	2	f	-0.818960	-0.832482	0.989473	0.666627	
135	2	g	-0.816188	-0.830510	0.990944	0.750372	
135	2	h	-0.001788	-0.002325	0.999999	1.582846	
135	2	j	-0.766710	-0.789706	0.977474	0.698917	
135	2	l	-0.001329	-0.001392	1.000000	0.479305	
135	2	m	-0.813643	-0.831974	0.988963	0.859979	
135	2	n	-0.826860	-0.824281	0.988871	-0.126964	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
142	1	a	-0.192139	-0.189532	0.999963	-1.337336	
142	1	d	-0.255597	-0.288940	0.940953	0.439755	
142	1	e	-0.000662	-0.000662	1.000000	0.000000	
142	1	h	-0.277455	-0.249826	0.914362	-0.302090	
142	1	j	-0.318605	-0.312101	0.862602	-0.057208	
142	1	m	-0.304107	-0.303750	0.940922	-0.004756	
142	1	n	-0.211922	-0.206682	0.972119	-0.098946	
142	2	a	-0.324310	-0.321402	0.999945	-1.265263	
142	2	b	-0.435620	-0.434607	0.999992	-1.209647	
142	2	c	-0.355980	-0.353568	0.999961	-1.251661	
142	2	d	-0.339456	-0.351042	0.962410	0.196452	
142	2	e	-0.000662	-0.000662	1.000000	0.000000	
142	2	f	-0.452453	-0.438539	0.964745	-0.255435	
142	2	g	-0.352275	-0.357367	0.961688	0.085868	
142	2	h	-0.277419	-0.054442	0.217178	-0.804482	
142	2	j	-0.316247	-0.360107	0.848330	0.370505	
142	2	l	-0.255325	-0.002208	0.398154	-1.037952	
142	2	m	-0.359307	-0.356064	0.955944	-0.051074	
142	2	n	-0.254817	-0.261328	0.978172	0.140630	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
158	1	a	0.038318	0.037274	0.999861	0.234405	
158	1	d	-0.162911	-0.142162	0.977468	-0.370037	
158	1	e	-0.001032	-0.001032	1.000000	0.000000	
158	1	h	-0.261133	-0.249736	0.993442	-0.384794	
158	1	j	-0.172847	-0.138819	0.969187	-0.519199	
158	1	m	-0.212637	-0.184153	0.975315	-0.489168	
158	1	n	-0.190370	-0.171699	0.972976	-0.305557	
158	2	a	0.090669	0.089328	0.999809	0.257478	
158	2	b	0.174475	0.173910	0.999972	0.285531	
158	2	c	0.116130	0.115103	0.999875	0.244216	
158	2	d	-0.113967	-0.096088	0.978496	-0.324375	
158	2	e	-0.001032	-0.001032	1.000000	0.000000	
158	2	f	0.016314	0.025684	0.981859	-0.184111	
158	2	g	-0.089306	-0.073014	0.977661	-0.289364	
158	2	h	-0.136551	-0.141136	0.993226	0.148835	
158	2	j	-0.103790	-0.098236	0.987446	-0.131824	
158	2	l	-0.007262	-0.010593	0.999166	0.305304	
158	2	m	-0.131206	-0.117746	0.983213	-0.277017	
158	2	n	-0.120778	-0.118780	0.981990	-0.039666	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
170	1	a	-0.716613	-0.716201	0.999987	-0.500158	
170	1	d	-0.731051	-0.740470	0.993749	0.536490	
170	1	e	-0.001109	-0.001109	1.000000	0.000000	
170	1	h	0.166612	0.161113	0.999923	1.938776	
170	1	j	-0.594562	-0.634013	0.987689	1.318577	
170	1	m	-0.636259	-0.653973	0.992768	0.822547	
170	1	n	-0.515708	-0.550422	0.992263	1.377149	
170	2	a	-0.678385	-0.678112	0.999979	-0.249519	
170	2	b	-0.637558	-0.637414	0.999996	-0.274505	
170	2	c	-0.678090	-0.677830	0.999983	-0.262757	
170	2	d	-0.691708	-0.716829	0.995616	1.497848	
170	2	e	-0.001109	-0.001109	1.000000	0.000000	
170	2	f	-0.696939	-0.729204	0.995410	1.803287	
170	2	g	-0.689511	-0.717631	0.995677	1.648823	
170	2	h	0.153761	0.148671	0.999951	2.235455	<****
170	2	j	-0.553947	-0.594241	0.994575	1.866306	
170	2	l	0.045678	0.044916	0.999997	1.372835	
170	2	m	-0.623943	-0.655047	0.995500	1.682713	
170	2	n	-0.533728	-0.577632	0.994568	1.987606	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
180	1	a	-0.297536	-0.299409	0.999723	0.416233	
180	1	d	-0.475084	-0.465962	0.972956	-0.222534	
180	1	e	-0.484654	-0.448249	0.943001	-0.609328	
180	1	h	-0.234297	-0.273204	0.994010	1.815255	
180	1	j	-0.395090	-0.383178	0.972203	-0.274375	
180	1	m	-0.473372	-0.462677	0.970381	-0.248950	
180	1	n	-0.428501	-0.421864	0.971957	-0.155029	
180	2	a	-0.219548	-0.222590	0.999568	0.530523	
180	2	b	-0.157462	-0.159036	0.999904	0.573887	
180	2	c	-0.213261	-0.215989	0.999684	0.554789	
180	2	d	-0.475058	-0.465784	0.965917	-0.201648	
180	2	e	-0.434632	-0.423526	0.925996	-0.160429	
180	2	f	-0.393963	-0.397721	0.966582	0.079270	
180	2	g	-0.468597	-0.462149	0.965305	-0.138573	
180	2	h	-0.352302	-0.400961	0.981130	1.333672	
180	2	j	-0.449142	-0.425860	0.976790	-0.599360	
180	2	l	-0.126480	-0.150684	0.997611	1.761979	
180	2	m	-0.476596	-0.459917	0.969177	-0.380275	
180	2	n	-0.445123	-0.424273	0.974688	-0.513802	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
182	1	a	-0.194413	-0.192936	0.999984	-1.386175	
182	1	d	-0.221779	-0.187207	0.993105	-1.583342	
182	1	e	-0.000462	-0.000462	1.000000	0.000000	
182	1	h	0.185728	0.201367	0.997233	-1.131017	
182	1	j	-0.008062	0.002560	0.992515	-0.459382	
182	1	m	-0.187365	-0.154416	0.988405	-1.160081	
182	1	n	-0.166005	-0.135654	0.982471	-0.867261	
182	2	a	-0.174304	-0.172579	0.999979	-1.414123	
182	2	b	-0.101067	-0.100631	0.999999	-1.424118	
182	2	c	-0.137425	-0.136297	0.999991	-1.417226	
182	2	d	-0.223362	-0.177736	0.991122	-1.837508	
182	2	e	-0.037327	-0.037327	1.000000	0.000000	
182	2	f	-0.147249	-0.126453	0.993692	-0.988232	
182	2	g	-0.195747	-0.158885	0.991635	-1.527192	
182	2	h	0.188673	0.198303	0.999043	-1.183623	
182	2	j	-0.009834	0.012329	0.987399	-0.738875	
182	2	l	0.157129	0.162794	0.999685	-1.208232	
182	2	m	-0.198272	-0.152594	0.987349	-1.538783	
182	2	n	-0.190478	-0.151898	0.984652	-1.180870	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
184	1	a	-0.132394	-0.132581	0.999986	0.160857	
184	1	d	0.129099	0.066967	0.964340	1.046246	
184	1	e	0.064221	0.004749	0.965202	1.010003	
184	1	h	-0.045491	-0.042961	0.999755	-0.511292	
184	1	j	0.003093	-0.010500	0.963262	0.224291	
184	1	m	0.080914	0.039387	0.962558	0.680278	
184	1	n	-0.007330	-0.045538	0.971267	0.713418	
184	2	a	-0.148937	-0.149158	0.999987	0.195327	
184	2	b	-0.139155	-0.139231	0.999998	0.166794	
184	2	c	-0.145872	-0.146048	0.999990	0.173707	
184	2	d	0.128770	0.054595	0.949794	1.052771	
184	2	e	0.142193	0.082934	0.958561	0.927009	
184	2	f	0.039625	-0.025722	0.948904	0.915649	
184	2	g	0.121064	0.044100	0.947888	1.071616	
184	2	h	-0.032087	-0.032241	0.999998	0.322612	
184	2	j	0.038815	-0.012448	0.957412	0.786357	
184	2	l	-0.004347	-0.004342	0.999999	-0.017192	
184	2	m	0.076227	0.015787	0.952486	0.878859	
184	2	n	-0.024803	-0.069690	0.970039	0.821506	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.146761	0.988414	0.148595	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.119714	0.995724	-0.125785	
203	1	j	0.072857	0.069456	0.988791	0.099277	
203	1	m	-0.097276	-0.103047	0.988140	0.164159	
203	1	n	-0.012827	-0.014390	0.994124	0.062847	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.053172	0.986919	-0.463065	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.021549	0.991421	-0.489875	
203	2	g	-0.070989	-0.053954	0.987333	-0.467457	
203	2	h	0.069005	0.074765	0.998039	-0.401902	
203	2	j	0.034705	0.046557	0.986288	-0.312227	
203	2	l	0.002329	0.002661	0.999996	-0.498452	
203	2	m	-0.044468	-0.029961	0.985690	-0.374055	
203	2	n	0.004985	0.015115	0.992221	-0.354036	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	0.008881	0.959878	-0.235964	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.204871	0.999818	2.116904	(****
207	1	j	0.168212	0.150318	0.997112	1.064410	
207	1	m	0.099118	0.097251	0.980516	0.042517	
207	1	n	0.064855	0.072536	0.979105	-0.168435	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.053102	0.967664	-0.548712	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	0.031567	0.953723	-1.058178	
207	2	g	0.018805	0.058703	0.966936	-0.694770	
207	2	h	0.202720	0.194292	0.999825	2.029956	(****
207	2	j	0.193024	0.189561	0.999021	0.356631	
207	2	l	0.175987	0.172265	0.999885	1.113286	
207	2	m	0.121846	0.138912	0.985794	-0.456642	
207	2	n	0.098820	0.122376	0.984216	-0.596537	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.642554	0.994251	0.678550	
212	1	e	-0.281644	-0.387175	0.968010	2.111565	<****
212	1	h	0.089775	0.084396	0.988412	0.173772	
212	1	j	-0.592539	-0.578060	0.978164	-0.418296	
212	1	m	-0.663788	-0.651190	0.994946	-0.801316	
212	1	n	-0.701485	-0.653017	0.988907	-1.929199	
212	2	a	-0.553674	-0.556193	0.999356	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.628426	0.947118	-1.070159	
212	2	e	-0.288411	-0.387069	0.975143	2.230195	<****
212	2	f	-0.681729	-0.602285	0.937197	-1.380904	
212	2	g	-0.692679	-0.630822	0.943921	-1.176219	
212	2	h	0.055268	0.047928	0.991705	0.279549	
212	2	j	-0.685346	-0.606738	0.942597	-1.427929	
212	2	l	0.004343	0.003638	0.999915	0.264216	
212	2	m	-0.710908	-0.637601	0.941238	-1.363291	
212	2	n	-0.731150	-0.632344	0.935889	-1.717563	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.461884	0.984751	0.777714	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.743589	0.972328	0.142141	
219	1	j	-0.671656	-0.711695	0.982453	1.120865	
219	1	m	-0.541866	-0.580514	0.983723	1.006440	
219	1	n	-0.562968	-0.600416	0.981076	0.923600	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.427720	0.981014	0.720933	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.413374	0.975909	0.647394	
219	2	g	-0.404803	-0.437486	0.980249	0.720044	
219	2	h	-0.769917	-0.747135	0.963845	-0.519063	
219	2	j	-0.594641	-0.640518	0.981724	1.164429	
219	2	l	-0.616435	-0.650398	0.977838	0.816354	
219	2	m	-0.504693	-0.549308	0.982342	1.085509	
219	2	n	-0.520418	-0.562119	0.979712	0.962351	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	-0.091837	0.953991	2.082487	(****
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.037803	0.999912	0.814601	
221	1	j	-0.072548	-0.179923	0.980257	2.031937	(****
221	1	m	0.012463	-0.140895	0.967212	2.257995	(****
221	1	n	-0.082838	-0.191504	0.987330	2.554274	(****
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	-0.090963	0.956845	1.690999	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.088283	0.971130	1.319809	
221	2	g	0.035177	-0.091772	0.957515	1.639048	
221	2	h	-0.034210	-0.036675	0.999963	1.076885	
221	2	j	-0.066140	-0.147974	0.983964	1.717527	
221	2	l	-0.001531	-0.001678	1.000000	1.399711	
221	2	m	-0.013498	-0.114270	0.975392	1.707021	
221	2	n	-0.092854	-0.156450	0.989795	1.673295	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.374690	0.948363	-0.490593	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.182934	0.999813	0.120842	
222	1	j	-0.403934	-0.376705	0.938304	-0.377074	
222	1	m	-0.430564	-0.399497	0.942107	-0.449155	
222	1	n	-0.194080	-0.160212	0.975691	-0.697601	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.391805	0.980736	-0.472869	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.336722	0.994392	-0.095047	
222	2	g	-0.410581	-0.393432	0.987102	-0.520168	
222	2	h	-0.022308	-0.021811	0.999995	-0.702709	
222	2	j	-0.430670	-0.407286	0.972594	-0.491216	
222	2	l	-0.002454	-0.002434	1.000000	-0.223601	
222	2	m	-0.442443	-0.421587	0.979479	-0.509322	
222	2	n	-0.256865	-0.223773	0.984624	-0.867158	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.677763	0.997143	-1.092917	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.373031	1.000000	-0.138144	
223	1	j	-0.422509	-0.420913	0.999851	-0.380351	
223	1	m	-0.695676	-0.686277	0.997616	-0.687806	
223	1	n	-0.639829	-0.629727	0.997056	-0.626973	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.707707	0.986766	-0.790231	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.699728	0.971249	-0.209693	
223	2	g	-0.731842	-0.706009	0.983403	-0.743238	
223	2	h	-0.361158	-0.361037	1.000000	-0.982439	
223	2	j	-0.389875	-0.391240	0.999863	0.334508	
223	2	l	-0.315803	-0.315725	1.000000	-1.159576	
223	2	m	-0.726692	-0.732142	0.988856	0.199866	
223	2	n	-0.670259	-0.678302	0.989974	0.287234	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.366376	0.951639	0.690588	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.338388	0.990355	-0.170894	
227	1	j	0.059223	-0.078257	0.942847	1.830749	
227	1	m	-0.227812	-0.318183	0.947080	1.284844	
227	1	n	-0.061523	-0.200120	0.913762	1.512095	
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.300352	0.955781	0.191304	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.301705	0.952469	0.294832	
227	2	g	-0.290104	-0.302799	0.952048	0.192151	
227	2	h	0.331148	0.335086	0.998555	-0.347107	
227	2	j	0.023736	-0.053345	0.976768	1.602716	
227	2	l	0.037912	0.085841	0.989605	-1.489756	
227	2	m	-0.214012	-0.252270	0.966465	0.678814	
227	2	n	-0.066552	-0.146941	0.962811	1.327089	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	d	-0.251948	-0.230046	0.986592	-0.599926	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.078429	0.999973	-0.981999	
230	1	j	0.058008	0.063166	0.998266	-0.382513	
230	1	m	-0.102239	-0.079651	0.968921	-0.396626	
230	1	n	-0.169797	-0.126726	0.970358	-0.779544	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.192806	0.998371	-0.367888	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.148999	0.999690	-0.099939	
230	2	g	-0.196159	-0.191492	0.998626	-0.395459	
230	2	h	0.064555	0.065137	0.999985	-0.465699	
230	2	j	0.074391	0.072703	0.999321	0.200236	
230	2	l	0.073664	0.074166	0.999994	-0.609238	
230	2	m	-0.075132	-0.070767	0.996823	-0.239286	
230	2	n	-0.137167	-0.125737	0.997244	-0.676492	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	d	-0.329375	-0.389761	0.987240	1.669537	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.132300	0.961816	0.591290	
235	1	j	-0.506407	-0.493855	0.990989	-0.456492	
235	1	m	-0.397157	-0.449965	0.983291	1.320095	
235	1	n	-0.457890	-0.489346	0.984272	0.844257	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.326042	0.988291	-1.043553	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.345244	0.976975	-1.406890	
235	2	g	-0.362648	-0.319956	0.986514	-1.159334	
235	2	h	-0.126989	-0.146510	0.991070	0.625347	
235	2	j	-0.453489	-0.412105	0.983240	-1.047251	
235	2	l	-0.014128	-0.016507	0.999903	0.724916	
235	2	m	-0.403257	-0.357926	0.984402	-1.160001	
235	2	n	-0.434261	-0.385742	0.982510	-1.184994	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.289763	0.912432	-0.984547	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209637	1.000000	-0.574103	
248	1	j	-0.240205	-0.226485	0.997492	-0.888240	
248	1	m	-0.384995	-0.311930	0.952670	-1.124287	
248	1	n	-0.315971	-0.261260	0.969052	-1.022399	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.290364	0.926808	-1.213387	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.297756	0.925675	-1.106347	
248	2	g	-0.385112	-0.283832	0.923286	-1.219670	
248	2	h	-0.209574	-0.209509	1.000000	-0.816489	
248	2	j	-0.253679	-0.239191	0.997948	-1.038575	
248	2	l	-0.206419	-0.206391	1.000000	-0.350382	
248	2	m	-0.426066	-0.346399	0.953436	-1.249369	
248	2	n	-0.348954	-0.287590	0.969677	-1.166073	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

CENTRAL PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.139596	0.974837	0.128531	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.201717	0.996858	-0.765583	
252	1	j	0.209856	0.226758	0.972531	-0.313510	
252	1	m	0.200105	0.199040	0.972810	0.019795	
252	1	n	0.253456	0.261383	0.975209	-0.156356	
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	d	0.188045	0.218636	0.984463	-0.750896	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.154452	0.983097	-1.379562	
252	2	g	0.196328	0.233146	0.983441	-0.876884	
252	2	h	0.024025	0.025611	0.999946	-0.645404	
252	2	j	0.244992	0.269456	0.991456	-0.818913	
252	2	l	0.000861	0.000946	1.000000	-0.415749	
252	2	m	0.248129	0.275464	0.988087	-0.776200	
252	2	n	0.295229	0.309959	0.993682	-0.581924	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.160415	0.995277	0.851457	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.104833	0.999526	1.737331	
203	1	j	0.072857	0.094659	0.993378	-0.828564	
203	1	m	-0.097276	-0.112616	0.993226	0.577519	
203	1	n	-0.012827	-0.012665	0.992025	-0.005580	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.081952	0.997292	0.690755	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.050291	0.997751	0.912521	
203	2	g	-0.070989	-0.083169	0.997262	0.719449	
203	2	h	0.069005	0.065814	0.999924	1.129174	
203	2	j	0.034705	0.026458	0.997600	0.519096	
203	2	l	0.002329	0.002250	1.000000	0.422963	
203	2	m	-0.044468	-0.059333	0.996541	0.780132	
203	2	n	0.004985	-0.003015	0.996376	0.409634	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	0.016294	0.972406	-0.425710	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.214531	0.999991	-0.758070	
207	1	j	0.168212	0.160671	0.999335	0.935811	
207	1	m	0.099118	0.099707	0.993702	-0.023563	
207	1	n	0.064855	0.063666	0.991169	0.040086	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.057765	0.975297	-0.721826	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	0.018031	0.959633	-0.918990	
207	2	g	0.018805	0.058379	0.974192	-0.779983	
207	2	h	0.202720	0.203888	0.999983	-0.909808	
207	2	j	0.193024	0.193242	0.999661	-0.038158	
207	2	l	0.175987	0.179723	0.999665	-0.655097	
207	2	m	0.121846	0.137139	0.994319	-0.646824	
207	2	n	0.098820	0.117626	0.993677	-0.752024	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.609182	0.990609	-0.972057	
212	1	e	-0.281644	-0.367326	0.976589	2.003152	(****
212	1	h	0.089775	0.092420	0.999949	-1.286973	
212	1	j	-0.592539	-0.564368	0.994148	-1.491171	
212	1	m	-0.663788	-0.641139	0.992193	-1.133026	
212	1	n	-0.701485	-0.686904	0.996452	-1.132692	
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.609941	0.968614	-1.737246	
212	2	e	-0.288411	-0.408138	0.969691	2.440264	(****
212	2	f	-0.681729	-0.606058	0.960850	-1.625723	
212	2	g	-0.692679	-0.616398	0.966042	-1.751584	
212	2	h	0.055268	0.056689	0.999967	-0.853978	
212	2	j	-0.685346	-0.640357	0.981874	-1.459850	
212	2	l	0.004343	0.004507	1.000000	-0.820542	
212	2	m	-0.710908	-0.652988	0.975446	-1.623423	
212	2	n	-0.731150	-0.700484	0.990188	-1.435622	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.428167	0.962522	-0.049107	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.724187	0.959347	-0.286572	
219	1	j	-0.671656	-0.654470	0.980216	-0.459517	
219	1	m	-0.541866	-0.537656	0.971903	-0.084605	
219	1	n	-0.562968	-0.551813	0.974578	-0.238642	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.379081	0.959384	-0.250061	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.357337	0.956167	-0.338536	
219	2	g	-0.404803	-0.387454	0.959993	-0.267448	
219	2	h	-0.769917	-0.742118	0.958057	-0.584331	
219	2	j	-0.594641	-0.604725	0.974991	0.225788	
219	2	l	-0.616435	-0.637544	0.973718	0.471012	
219	2	m	-0.504693	-0.499509	0.970628	-0.099139	
219	2	n	-0.520418	-0.514222	0.972638	-0.124054	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	-0.001118	0.976299	1.318525	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.042102	0.999526	0.874369	
221	1	j	-0.072548	-0.111765	0.965440	0.560721	
221	1	m	0.012463	-0.050746	0.972285	1.006163	
221	1	n	-0.082838	-0.124017	0.973413	0.671893	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	-0.042582	0.966224	1.206113	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.069580	0.970821	1.021014	
221	2	g	0.035177	-0.048160	0.964882	1.179566	
221	2	h	-0.034210	-0.038443	0.999778	0.751501	
221	2	j	-0.066140	-0.115411	0.969342	0.747931	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.079919	0.967568	0.978256	
221	2	n	-0.092854	-0.135324	0.973427	0.693911	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.362235	0.966419	-0.834865	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.180458	0.999724	-0.379519	
222	1	j	-0.403934	-0.384673	0.986290	-0.564383	
222	1	m	-0.430564	-0.392961	0.971649	-0.770674	
222	1	n	-0.194080	-0.191052	0.990843	-0.101972	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.366092	0.971400	-0.901713	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.291939	0.956741	-0.749853	
222	2	g	-0.410581	-0.363829	0.966927	-0.876184	
222	2	h	-0.022308	-0.022041	0.999997	-0.529832	
222	2	j	-0.430670	-0.395590	0.978939	-0.833251	
222	2	l	-0.002454	-0.002434	1.000000	-0.223601	
222	2	m	-0.442443	-0.399871	0.971535	-0.872841	
222	2	n	-0.256865	-0.242209	0.989115	-0.458413	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.673473	0.991966	-0.836911	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.373048	1.000000	0.364864	
223	1	j	-0.422509	-0.421755	0.999917	-0.241072	
223	1	m	-0.695676	-0.674331	0.994234	-0.972685	
223	1	n	-0.639829	-0.612636	0.993787	-1.110079	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.670997	0.978573	-1.397333	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.631819	0.972828	-1.497024	
223	2	g	-0.731842	-0.667450	0.977200	-1.415382	
223	2	h	-0.361158	-0.361167	1.000000	0.326963	
223	2	j	-0.389875	-0.391874	0.999818	0.424475	
223	2	l	-0.315803	-0.315787	1.000000	-0.442164	
223	2	m	-0.726692	-0.696036	0.982395	-0.840189	
223	2	n	-0.670259	-0.634363	0.980962	-0.879345	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.282110	0.914091	-0.442966	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.339230	0.990074	-0.196856	
227	1	j	0.059223	0.091133	0.963962	-0.533222	
227	1	m	-0.227812	-0.182803	0.923801	-0.527222	
227	1	n	-0.061523	-0.058520	0.942958	-0.039835	
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.207730	0.891892	-0.800005	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.193110	0.881723	-0.845620	
227	2	g	-0.290104	-0.201262	0.880707	-0.840625	
227	2	h	0.331148	0.336789	0.997307	-0.364300	
227	2	j	0.023736	0.066934	0.941332	-0.564943	
227	2	l	0.037912	0.094829	0.985476	-1.497451	
227	2	m	-0.214012	-0.140513	0.898157	-0.741197	
227	2	n	-0.066552	-0.041382	0.925626	-0.292365	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	d	-0.251948	-0.193763	0.992150	-2.044762	(****
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.083894	0.999782	-1.492281	
230	1	j	0.058008	0.085373	0.997907	-1.846177	
230	1	m	-0.102239	-0.033802	0.989282	-2.042462	(****
230	1	n	-0.169797	-0.093877	0.986758	-2.043883	(****
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.153679	0.995213	-1.966987	(****
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.120400	0.997755	-1.905176	
230	2	g	-0.196159	-0.153784	0.995524	-1.965533	(****
230	2	h	0.064555	0.072050	0.999776	-1.545534	
230	2	j	0.074391	0.091716	0.999212	-1.905873	
230	2	l	0.073664	0.077823	0.999931	-1.548143	
230	2	m	-0.075132	-0.026887	0.994336	-1.978446	(****
230	2	n	-0.137167	-0.087740	0.994160	-2.000038	(****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	d	-0.329375	-0.327040	0.987682	-0.066835	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.247013	0.946995	2.020411	(****
235	1	j	-0.506407	-0.510119	0.991283	0.138534	
235	1	m	-0.397157	-0.402545	0.989366	0.171048	
235	1	n	-0.457890	-0.471652	0.991153	0.493903	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.362743	0.989625	0.031704	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.426475	0.987749	0.415838	
235	2	g	-0.362648	-0.366155	0.989273	0.109119	
235	2	h	-0.126989	-0.144205	0.998618	1.398178	
235	2	j	-0.453489	-0.464915	0.992450	0.442774	
235	2	l	-0.014128	-0.015324	0.999989	1.097803	
235	2	m	-0.403257	-0.412923	0.990933	0.333222	
235	2	n	-0.434261	-0.452864	0.992461	0.711631	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.471730	0.951889	1.475685	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209658	1.000000	-0.570606	
248	1	j	-0.240205	-0.265625	0.998324	1.990975	<****
248	1	m	-0.384995	-0.462279	0.980726	1.857652	
248	1	n	-0.315971	-0.383744	0.986235	1.889339	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.445397	0.946089	0.846063	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.420513	0.911508	0.381212	
248	2	g	-0.385112	-0.439403	0.937375	0.750970	
248	2	h	-0.209574	-0.209548	1.000000	-0.414941	
248	2	j	-0.253679	-0.265486	0.999056	1.248458	
248	2	l	-0.206419	-0.206393	1.000000	-0.588891	
248	2	m	-0.426066	-0.473893	0.972654	1.010930	
248	2	n	-0.348954	-0.399141	0.985413	1.390164	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL DEMONSTRATIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.143051	0.993549	0.123453	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.183418	0.997165	0.243299	
252	1	j	0.209856	0.227944	0.991894	-0.616895	
252	1	m	0.200105	0.204179	0.996458	-0.209669	
252	1	n	0.253456	0.257340	0.998579	-0.319601	
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	d	0.188045	0.209631	0.991000	-0.695632	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.136398	0.978952	-0.860338	
252	2	g	0.196328	0.221539	0.988708	-0.726679	
252	2	h	0.024025	0.025375	0.999978	-0.867120	
252	2	j	0.244992	0.263321	0.995845	-0.878591	
252	2	l	0.000861	0.000931	1.000000	-0.374169	
252	2	m	0.248129	0.267009	0.993097	-0.703734	
252	2	n	0.295229	0.307327	0.996937	-0.685320	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.164897	0.998851	2.123474	(****
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.112489	0.998845	0.417553	
203	1	j	0.072857	0.044487	0.997839	1.882983	
203	1	m	-0.097276	-0.120785	0.998426	1.831821	
203	1	n	-0.012827	-0.034324	0.997144	1.240257	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.087550	0.998257	1.275061	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.046142	0.998636	0.825071	
203	2	g	-0.070989	-0.087114	0.998248	1.190405	
203	2	h	0.069005	0.066168	0.999664	0.477835	
203	2	j	0.034705	0.016007	0.997727	1.209191	
203	2	l	0.002329	0.002293	0.999999	0.130191	
203	2	m	-0.044468	-0.063126	0.997800	1.227605	
203	2	n	0.004985	-0.015586	0.996083	1.013236	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	-0.002993	0.985863	-0.081692	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.213839	1.000000	-0.149109	
207	1	j	0.168212	0.168826	0.999672	-0.108723	
207	1	m	0.099118	0.100167	0.994839	-0.046379	
207	1	n	0.064855	0.066004	0.995427	-0.053844	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.024626	0.983560	-0.066469	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.032108	0.975500	-0.165235	
207	2	g	0.018805	0.021785	0.982335	-0.070929	
207	2	h	0.202720	0.202758	1.000000	-0.282227	
207	2	j	0.193024	0.193883	0.999780	-0.186594	
207	2	l	0.175987	0.176014	1.000000	-0.304390	
207	2	m	0.121846	0.123271	0.994608	-0.061867	
207	2	n	0.098820	0.101466	0.995993	-0.132850	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.623503	0.993079	-0.398576	
212	1	e	-0.281644	-0.357325	0.981454	1.985336	<****
212	1	h	0.089775	0.095909	0.999810	-1.546373	
212	1	j	-0.592539	-0.567812	0.988673	-0.970118	
212	1	m	-0.663788	-0.654131	0.993403	-0.543840	
212	1	n	-0.701485	-0.693472	0.994068	-0.499836	
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.658219	0.988943	-1.072763	
212	2	e	-0.288411	-0.364032	0.988204	2.453464	<****
212	2	f	-0.681729	-0.668706	0.987318	-0.540059	
212	2	g	-0.692679	-0.668444	0.988268	-1.028006	
212	2	h	0.055268	0.057284	0.999917	-0.769452	
212	2	j	-0.685346	-0.673404	0.992149	-0.630063	
212	2	l	0.004343	0.004398	1.000000	-0.496427	
212	2	m	-0.710908	-0.694421	0.991103	-0.834453	
212	2	n	-0.731150	-0.724208	0.995448	-0.515767	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.498884	0.988495	1.872636	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.754061	0.977807	0.455770	
219	1	j	-0.671656	-0.710030	0.993201	1.602030	
219	1	m	-0.541866	-0.583768	0.994165	1.709219	
219	1	n	-0.562968	-0.602841	0.995046	1.767546	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.487634	0.984689	2.149739	(****
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.476709	0.982824	2.120405	(****
219	2	g	-0.404803	-0.497617	0.984036	2.127564	(****
219	2	h	-0.769917	-0.781381	0.977978	0.346766	
219	2	j	-0.594641	-0.644089	0.993171	1.869720	
219	2	l	-0.616435	-0.661971	0.985519	1.303017	
219	2	m	-0.504693	-0.567119	0.991819	2.041780	(****
219	2	n	-0.520418	-0.575407	0.994148	2.108793	(****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	0.023421	0.986470	1.184675	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.042933	0.999682	1.190980	
221	1	j	-0.072548	-0.111107	0.987392	0.912237	
221	1	m	0.012463	-0.033502	0.986504	1.047693	
221	1	n	-0.082838	-0.117352	0.988757	0.865258	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	-0.023433	0.981267	1.247183	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.059320	0.985956	1.241323	
221	2	g	0.035177	-0.030448	0.980937	1.259488	
221	2	h	-0.034210	-0.039901	0.999854	1.244787	
221	2	j	-0.066140	-0.107814	0.987380	0.985108	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.065346	0.985043	1.123404	
221	2	n	-0.092854	-0.128436	0.989195	0.910719	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.425795	0.994763	0.887940	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.182851	0.999987	0.385119	
222	1	j	-0.403934	-0.412879	0.994999	0.437311	
222	1	m	-0.430564	-0.443776	0.995146	0.662838	
222	1	n	-0.194080	-0.195484	0.998009	0.101443	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.423575	0.996817	0.779034	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.353499	0.996737	0.859394	
222	2	g	-0.410581	-0.423674	0.996720	0.789160	
222	2	h	-0.022308	-0.022367	1.000000	0.407233	
222	2	j	-0.430670	-0.438467	0.997046	0.501973	
222	2	l	-0.002454	-0.002558	1.000000	1.162750	
222	2	m	-0.442443	-0.453278	0.996697	0.662604	
222	2	n	-0.256865	-0.256084	0.999213	-0.091133	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.698793	0.997560	0.291609	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.372969	1.000000	-0.461154	
223	1	j	-0.422509	-0.421740	0.999847	-0.181020	
223	1	m	-0.695676	-0.702973	0.996926	0.481499	
223	1	n	-0.639829	-0.657796	0.993868	0.776773	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.737229	0.997075	0.351518	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.718622	0.997960	0.761672	
223	2	g	-0.731842	-0.736598	0.997224	0.349586	
223	2	h	-0.361158	-0.360504	0.999993	-0.705861	
223	2	j	-0.389875	-0.389767	0.999917	-0.034085	
223	2	l	-0.315803	-0.315534	0.999999	-0.862837	
223	2	m	-0.726692	-0.735723	0.997116	0.637447	
223	2	n	-0.670259	-0.688063	0.994686	0.849540	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.333884	0.967730	0.236489	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.337967	0.990109	-0.154540	
227	1	j	0.059223	0.032169	0.994746	1.181649	
227	1	m	-0.227812	-0.256474	0.983208	0.719669	
227	1	n	-0.061523	-0.100363	0.990614	1.271679	
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.304077	0.953233	0.243191	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.309691	0.959723	0.452175	
227	2	g	-0.290104	-0.307485	0.949699	0.257034	
227	2	h	0.331148	0.336421	0.997302	-0.340260	
227	2	j	0.023736	0.009103	0.987400	0.412338	
227	2	l	0.037912	0.094536	0.985449	-1.488346	
227	2	m	-0.214012	-0.238251	0.966808	0.431890	
227	2	n	-0.066552	-0.097547	0.986636	0.850753	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

Q	S	TW	Correlation Coefficients r_{ju}	r_{jr}	r_{ur}	Significance Level Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	d	-0.251948	-0.206694	0.996406	-2.337259	<****
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.078775	0.999878	-0.562023	
230	1	j	0.058008	0.060077	0.999310	-0.243284	
230	1	m	-0.102239	-0.081183	0.993916	-0.835304	
230	1	n	-0.169797	-0.154411	0.998159	-1.117259	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.140121	0.995011	-2.508480	<****
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.100839	0.996083	-2.403394	<****
230	2	g	-0.196159	-0.139163	0.994956	-2.477595	<****
230	2	h	0.064555	0.067459	0.999761	-0.579694	
230	2	j	0.074391	0.085987	0.995579	-0.539230	
230	2	l	0.073664	0.077534	0.999564	-0.572658	
230	2	m	-0.075132	-0.032329	0.993257	-1.609020	
230	2	n	-0.137167	-0.086693	0.995136	-2.235707	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	d	-0.329375	-0.399434	0.986186	1.850739	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.129838	0.987257	0.956826	
235	1	j	-0.506407	-0.541177	0.975213	0.769463	
235	1	m	-0.397157	-0.460244	0.982322	1.522810	
235	1	n	-0.457890	-0.509970	0.980055	1.229786	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.419628	0.987110	1.609210	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.466748	0.986055	1.481726	
235	2	g	-0.362648	-0.420482	0.986986	1.601014	
235	2	h	-0.126989	-0.137956	0.998150	0.771022	
235	2	j	-0.453489	-0.509340	0.981978	1.374888	
235	2	l	-0.014128	-0.015413	0.999980	0.860525	
235	2	m	-0.403257	-0.465795	0.985032	1.633771	
235	2	n	-0.434261	-0.494363	0.984853	1.582275	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.337026	0.940801	-0.553664	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209649	1.000000	-1.213183	
248	1	j	-0.240205	-0.233500	0.998365	-0.539117	
248	1	m	-0.384995	-0.360608	0.978155	-0.561105	
248	1	n	-0.315971	-0.296292	0.987527	-0.584208	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.366638	0.944120	-0.318389	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.349852	0.934141	-0.503549	
248	2	g	-0.385112	-0.362132	0.940619	-0.321994	
248	2	h	-0.209574	-0.209543	1.000000	-0.862734	
248	2	j	-0.253679	-0.247293	0.998910	-0.630726	
248	2	l	-0.206419	-0.206393	1.000000	-0.627168	
248	2	m	-0.426066	-0.414015	0.973895	-0.260055	
248	2	n	-0.348954	-0.337393	0.986337	-0.332816	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RELATIVE PRONOUNS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.261077	0.968427	-1.962495	(****
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.214784	0.997835	-1.766706	
252	1	j	0.209856	0.322890	0.974105	-2.137546	(****
252	1	m	0.200105	0.294093	0.979743	-2.009274	(****
252	1	n	0.253456	0.303011	0.991606	-1.660919	
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	d	0.188045	0.233418	0.989041	-1.322697	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.093920	0.993518	0.042944	
252	2	g	0.196328	0.231480	0.990525	-1.104512	
252	2	h	0.024025	0.028344	0.999957	-1.978265	(****
252	2	j	0.244992	0.297828	0.991662	-1.770730	
252	2	l	0.000861	0.001181	1.000000	-1.939500	
252	2	m	0.248129	0.288191	0.992227	-1.399396	
252	2	n	0.295229	0.316754	0.995733	-1.030039	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.122017	0.997337	-1.179376	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.121139	0.999748	-0.797060	
203	1	j	0.072857	0.097196	0.996786	-1.327002	
203	1	m	-0.097276	-0.075291	0.996669	-1.177638	
203	1	n	-0.012827	0.007192	0.997692	-1.284602	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.047237	0.996690	-1.238604	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.019334	0.997846	-1.124668	
203	2	g	-0.070989	-0.048683	0.996825	-1.222133	
203	2	h	0.069005	0.071266	0.999944	-0.931962	
203	2	j	0.034705	0.056862	0.996951	-1.237896	
203	2	l	0.002329	0.002487	1.000000	-1.196375	
203	2	m	-0.044468	-0.021596	0.996640	-1.216922	
203	2	n	0.004985	0.023496	0.998056	-1.294277	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	-0.015021	0.997702	0.590862	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.213813	1.000000	0.959972	
207	1	j	0.168212	0.166324	0.999935	0.752712	
207	1	m	0.099118	0.094709	0.999233	0.505890	
207	1	n	0.064855	0.063781	0.999268	0.125732	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.013465	0.998701	0.743144	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.045760	0.999586	0.852672	
207	2	g	0.018805	0.010348	0.998827	0.780997	
207	2	h	0.202720	0.202736	1.000000	-0.935803	
207	2	j	0.193024	0.192399	0.999987	0.565203	
207	2	l	0.175987	0.176003	1.000000	-0.695023	
207	2	m	0.121846	0.118406	0.999763	0.711351	
207	2	n	0.098820	0.098040	0.999844	0.198758	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	p > .05
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.646427	0.998339	1.592860	
212	1	e	-0.281644	-0.326587	0.984881	1.317287	
212	1	h	0.089775	0.049367	0.978870	0.965636	
212	1	j	-0.592539	-0.616077	0.997809	1.982061	(****
212	1	m	-0.663788	-0.677796	0.998814	1.732608	
212	1	n	-0.701485	-0.708855	0.999517	1.517207	
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.683279	0.997142	0.013529	
212	2	e	-0.288411	-0.320491	0.992082	1.299060	
212	2	f	-0.681729	-0.682627	0.996382	0.070769	
212	2	g	-0.692679	-0.692217	0.996861	-0.039660	
212	2	h	0.055268	0.031040	0.992915	0.998164	
212	2	j	-0.685346	-0.689855	0.999186	0.741864	
212	2	l	0.004343	0.002371	0.999954	1.011592	
212	2	m	-0.710908	-0.710451	0.998167	-0.052548	
212	2	n	-0.731150	-0.728244	0.999230	-0.525755	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.475251	0.995865	2.001406	<****
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.753976	0.977907	0.454360	
219	1	j	-0.671656	-0.711427	0.994462	1.776276	
219	1	m	-0.541866	-0.581991	0.995766	1.877988	
219	1	n	-0.562968	-0.601222	0.996420	1.941727	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.448579	0.993423	1.915945	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.435879	0.992508	1.870053	
219	2	g	-0.404803	-0.458173	0.993070	1.884412	
219	2	h	-0.769917	-0.781529	0.978468	0.355104	
219	2	j	-0.594641	-0.640988	0.993603	1.823538	
219	2	l	-0.616435	-0.661934	0.985596	1.305042	
219	2	m	-0.504693	-0.553519	0.993948	1.889240	
219	2	n	-0.520418	-0.565979	0.995212	1.968895	<****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	0.062879	0.993247	0.404370	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.041861	0.998636	0.498106	
221	1	j	-0.072548	-0.078011	0.994438	0.194354	
221	1	m	0.012463	0.000665	0.991223	0.333200	
221	1	n	-0.082838	-0.087382	0.993039	0.144611	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	0.024147	0.987793	0.403540	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.020278	0.989501	0.425594	
221	2	g	0.035177	0.017615	0.987358	0.413433	
221	2	h	-0.034210	-0.040280	0.998956	0.497367	
221	2	j	-0.066140	-0.076664	0.988966	0.265752	
221	2	l	-0.001531	-0.001697	0.999999	0.524866	
221	2	m	-0.013498	-0.028967	0.987979	0.373397	
221	2	n	-0.092854	-0.102614	0.990053	0.260147	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.399368	0.992735	-0.314716	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.181549	0.999816	-0.206736	
222	1	j	-0.403934	-0.402929	0.999621	-0.178444	
222	1	m	-0.430564	-0.422595	0.992656	-0.324801	
222	1	n	-0.194080	-0.173806	0.990370	-0.664064	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.409847	0.996592	-0.059321	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.337874	0.995228	-0.046988	
222	2	g	-0.410581	-0.409606	0.996027	-0.053649	
222	2	h	-0.022308	-0.022249	0.999999	-0.223064	
222	2	j	-0.430670	-0.430466	0.997729	-0.014989	
222	2	l	-0.002454	-0.002454	1.000000	0.000000	
222	2	m	-0.442443	-0.441810	0.995928	-0.034932	
222	2	n	-0.256865	-0.234907	0.987643	-0.643495	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.731242	0.999994	-1.728851	
223	1	d	-0.694873	-0.688107	0.996236	-0.400950	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.373034	1.000000	-0.333290	
223	1	j	-0.422509	-0.417174	0.999865	-1.296866	
223	1	m	-0.695676	-0.674388	0.997288	-1.344104	
223	1	n	-0.639829	-0.616235	0.997039	-1.355832	
223	2	a	-0.754136	-0.752364	0.999987	-1.630948	
223	2	b	-0.718551	-0.717670	0.999996	-1.532998	
223	2	c	-0.752835	-0.751203	0.999989	-1.633879	
223	2	d	-0.732325	-0.731320	0.999396	-0.158490	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.708283	0.999766	-0.229166	
223	2	g	-0.731842	-0.730878	0.999493	-0.165860	
223	2	h	-0.361158	-0.361080	1.000000	-0.440584	
223	2	j	-0.389875	-0.387820	0.999958	-0.896217	
223	2	l	-0.315803	-0.315761	1.000000	-0.647434	
223	2	m	-0.726692	-0.719852	0.999597	-1.196002	
223	2	n	-0.670259	-0.662359	0.999485	-1.154575	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.351401	0.940715	0.416855	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.338163	0.990107	-0.161140	
227	1	j	0.059223	0.019964	0.945642	0.533209	
227	1	m	-0.227812	-0.253677	0.927851	0.314029	
227	1	n	-0.061523	-0.077847	0.928269	0.193250	
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.285871	0.950694	-0.034733	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.255137	0.948825	-0.394053	
227	2	g	-0.290104	-0.280970	0.947762	-0.132004	
227	2	h	0.331148	0.336708	0.997306	-0.359012	
227	2	j	0.023736	0.000551	0.954392	0.343405	
227	2	l	0.037912	0.094723	0.985467	-1.494170	
227	2	m	-0.214012	-0.211916	0.939348	-0.027571	
227	2	n	-0.066552	-0.070422	0.933953	0.047729	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373328	0.999999	-0.982605	
230	1	d	-0.251948	-0.251866	0.935136	-0.001026	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.077018	0.999997	-0.423936	
230	1	j	0.058008	0.054187	0.996789	0.208135	
230	1	m	-0.102239	-0.112390	0.968626	0.177683	
230	1	n	-0.169797	-0.175705	0.967702	0.102896	
230	2	a	-0.305620	-0.305273	0.999999	-1.108923	
230	2	b	-0.234657	-0.234507	1.000000	-1.228204	
230	2	c	-0.301274	-0.300957	0.999999	-1.117145	
230	2	d	-0.197532	-0.207270	0.930252	0.116136	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.162809	0.941879	0.171498	
230	2	g	-0.196159	-0.205976	0.925394	0.113172	
230	2	h	0.064555	0.064911	0.999999	-1.178203	
230	2	j	0.074391	0.069176	0.995794	0.248452	
230	2	l	0.073664	0.073596	0.999999	0.183226	
230	2	m	-0.075132	-0.089965	0.949348	0.203868	
230	2	n	-0.137167	-0.148929	0.945605	0.157109	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	d	-0.329375	-0.393828	0.989697	1.959819	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.104226	0.989185	0.295166	
235	1	j	-0.506407	-0.574452	0.983211	1.736605	
235	1	m	-0.397157	-0.467259	0.986296	1.884332	
235	1	n	-0.457890	-0.525953	0.984831	1.783855	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.420046	0.989110	1.751795	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.475978	0.984102	1.615723	
235	2	g	-0.362648	-0.421563	0.988388	1.717625	
235	2	h	-0.126989	-0.142522	0.998556	1.234739	
235	2	j	-0.453489	-0.519651	0.983442	1.671045	
235	2	l	-0.014128	-0.015829	0.999985	1.316555	
235	2	m	-0.403257	-0.467173	0.986074	1.722661	
235	2	n	-0.434261	-0.495595	0.985837	1.661679	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.440045	0.965948	1.170448	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209766	1.000000	0.926858	
248	1	j	-0.240205	-0.258084	0.998240	1.379506	
248	1	m	-0.384995	-0.429285	0.982199	1.132802	
248	1	n	-0.315971	-0.365961	0.985698	1.384493	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.436862	0.957483	0.807662	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.438841	0.953213	0.815199	
248	2	g	-0.385112	-0.432598	0.954699	0.769845	
248	2	h	-0.209574	-0.209617	1.000000	0.913745	
248	2	j	-0.253679	-0.267125	0.998534	1.142641	
248	2	i	-0.206419	-0.206457	1.000000	0.483405	
248	2	m	-0.426066	-0.454715	0.975104	0.636954	
248	2	n	-0.348954	-0.385543	0.982370	0.929850	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

NOMINAL SUBSTITUTES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.170472	0.997623	-1.498251	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.192869	0.997087	-0.294817	
252	1	j	0.209856	0.243429	0.991116	-1.092112	
252	1	m	0.200105	0.224950	0.996364	-1.257715	
252	1	n	0.253456	0.267481	0.996404	-0.724663	
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	d	0.188045	0.204110	0.996135	-0.789158	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.108899	0.993730	-0.526705	
252	2	g	0.196328	0.211898	0.995575	-0.716091	
252	2	h	0.024025	0.025382	0.999952	-0.585217	
252	2	j	0.244992	0.260262	0.995688	-0.719136	
252	2	i	0.000861	0.000966	1.000000	-0.513527	
252	2	m	0.248129	0.261814	0.995253	-0.615062	
252	2	n	0.295229	0.301217	0.995104	-0.268839	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106373	0.999971	0.590049	
203	1	d	-0.141626	-0.140762	0.999953	-0.394116	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.117082	1.000000	-1.013372	
203	1	j	0.072857	0.074912	0.999952	-0.919138	
203	1	m	-0.097276	-0.095974	0.999947	-0.555197	
203	1	n	-0.012827	-0.010821	0.999942	-0.808327	
203	2	a	-0.042574	-0.043888	0.999949	0.565778	
203	2	b	0.005796	0.005309	0.999993	0.554977	
203	2	c	-0.037613	-0.038727	0.999963	0.568352	
203	2	d	-0.070321	-0.070567	0.999974	0.148748	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.036419	0.999995	0.224800	
203	2	g	-0.070989	-0.071214	0.999979	0.153097	
203	2	h	0.069005	0.069169	1.000000	-1.007216	
203	2	j	0.034705	0.035171	0.999979	-0.313427	
203	2	l	0.002329	0.002420	1.000000	-0.998119	
203	2	m	-0.044468	-0.044404	0.999976	-0.040565	
203	2	n	0.004985	0.005557	0.999975	-0.352308	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	-0.007045	0.999902	0.312993	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.213822	1.000000	0.000000	
207	1	j	0.168212	0.168215	1.000000	-0.016288	
207	1	m	0.099118	0.099005	0.999977	0.075470	
207	1	n	0.064855	0.065020	0.999980	-0.115978	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.022143	0.999743	-0.041548	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.041326	0.999955	0.491185	
207	2	g	0.018805	0.018842	0.999799	-0.008286	
207	2	h	0.202720	0.202736	1.000000	-1.053751	
207	2	j	0.193024	0.193329	0.999997	-0.597415	
207	2	l	0.175987	0.176003	1.000000	-0.695023	
207	2	m	0.121846	0.122559	0.999958	-0.350778	
207	2	n	0.098820	0.100521	0.999972	-1.017328	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.533515	0.999961	0.726140	
212	1	d	-0.630979	-0.641000	0.998405	1.092061	
212	1	e	-0.281644	-0.327318	0.984898	1.339204	
212	1	h	0.089775	0.089767	1.000000	0.305103	
212	1	j	-0.592539	-0.597680	0.999279	0.813148	
212	1	m	-0.663788	-0.670554	0.998968	0.954455	
212	1	n	-0.701485	-0.702827	0.999617	0.332593	
212	2	a	-0.553674	-0.554955	0.999949	0.738458	
212	2	b	-0.536118	-0.536919	0.999989	0.980276	
212	2	c	-0.560078	-0.561285	0.999962	0.807735	
212	2	d	-0.683127	-0.690872	0.998363	0.891347	
212	2	e	-0.288411	-0.320948	0.992084	1.317437	
212	2	f	-0.681729	-0.691854	0.998815	1.327136	
212	2	g	-0.692679	-0.700712	0.998380	0.938043	
212	2	h	0.055268	0.054862	0.999998	1.143936	
212	2	j	-0.685346	-0.690524	0.999308	0.916263	
212	2	l	0.004343	0.004343	1.000000	0.000000	
212	2	m	-0.710908	-0.716524	0.999002	0.858428	
212	2	n	-0.731150	-0.732718	0.999761	0.511488	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331558	0.999999	0.428381	
219	1	d	-0.431196	-0.430929	0.999993	-0.317224	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.738106	0.999999	0.491606	
219	1	j	-0.671656	-0.671557	0.999995	-0.176292	
219	1	m	-0.541866	-0.541748	0.999985	-0.102544	
219	1	n	-0.562968	-0.562841	0.999981	-0.098643	
219	2	a	-0.253123	-0.253282	0.999997	0.290265	
219	2	b	-0.223322	-0.223384	0.999999	0.222491	
219	2	c	-0.264359	-0.264516	0.999998	0.312450	
219	2	d	-0.395491	-0.395332	0.999997	-0.307938	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.380539	0.999999	-0.283615	
219	2	g	-0.404803	-0.404651	0.999998	-0.313673	
219	2	h	-0.769917	-0.770058	0.999996	0.327318	
219	2	j	-0.594641	-0.594619	0.999993	-0.029247	
219	2	l	-0.616435	-0.616571	0.999998	0.375072	
219	2	m	-0.504693	-0.504606	0.999995	-0.128464	
219	2	n	-0.520418	-0.520337	0.999995	-0.119573	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	0.068418	0.999671	1.021262	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.034912	1.000000	0.000000	
221	1	j	-0.072548	-0.078012	0.999855	1.204255	
221	1	m	0.012463	0.006072	0.999766	1.105415	
221	1	n	-0.082838	-0.085460	0.999962	1.131846	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	0.039240	0.999922	0.524472	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.004296	0.999986	0.347925	
221	2	g	0.035177	0.033652	0.999936	0.505707	
221	2	h	-0.034210	-0.034210	1.000000	0.000000	
221	2	j	-0.066140	-0.067458	0.999979	0.754856	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.014859	0.999966	0.617133	
221	2	n	-0.092854	-0.093602	0.999993	0.738241	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.126934	0.999988	-1.509196	
227	1	d	-0.321199	-0.397123	0.974481	1.576458	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.337955	0.990110	-0.154117	
227	1	j	0.059223	0.019363	0.992820	1.489250	
227	1	m	-0.227812	-0.300248	0.976206	1.524571	
227	1	n	-0.061523	-0.123219	0.980781	1.413555	
227	2	a	-0.122705	-0.120630	0.999981	-1.499219	
227	2	b	-0.133699	-0.132765	0.999996	-1.489035	
227	2	c	-0.126353	-0.124485	0.999984	-1.497910	
227	2	d	-0.288205	-0.354105	0.980924	1.566304	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.338312	0.982229	1.381232	
227	2	g	-0.290104	-0.356231	0.980555	1.557834	
227	2	h	0.331148	0.336407	0.997310	-0.339819	
227	2	j	0.023736	-0.019330	0.991257	1.457485	
227	2	l	0.037912	0.094589	0.985453	-1.489917	
227	2	m	-0.214012	-0.277478	0.980964	1.488277	
227	2	n	-0.066552	-0.125085	0.982356	1.399863	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	e
230	1	d	-0.251948	-0.251948	1.000000	0.000000	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.076769	1.000000	0.000000	
230	1	j	0.058008	0.058008	1.000000	0.000000	
230	1	m	-0.102239	-0.102239	1.000000	0.000000	
230	1	n	-0.169797	-0.169797	1.000000	0.000000	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.197532	1.000000	0.000000	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.149564	1.000000	0.000000	
230	2	g	-0.196159	-0.196159	1.000000	0.000000	
230	2	h	0.064555	0.064555	1.000000	0.000000	
230	2	j	0.074391	0.074391	1.000000	0.000000	
230	2	l	0.073664	0.073664	1.000000	0.000000	
230	2	m	-0.075132	-0.075132	1.000000	0.000000	
230	2	n	-0.137167	-0.137166	1.000000	-0.468735	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.029072	0.999980	-1.384357	
235	1	d	-0.329375	-0.329837	0.999961	0.235258	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.094042	1.000000	-0.219935	
235	1	j	-0.506407	-0.505755	0.999931	-0.271898	
235	1	m	-0.397157	-0.396960	0.999949	-0.090133	
235	1	n	-0.457890	-0.456801	0.999928	-0.432790	
235	2	a	-0.160684	-0.157446	0.999947	-1.343212	
235	2	b	-0.226216	-0.224573	0.999986	-1.320368	
235	2	c	-0.155708	-0.152832	0.999958	-1.342845	
235	2	d	-0.361740	-0.361667	0.999977	-0.049459	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.412476	0.999993	-0.056523	
235	2	g	-0.362648	-0.362592	0.999981	-0.041941	
235	2	h	-0.126989	-0.126878	1.000000	-0.649706	
235	2	j	-0.453489	-0.453002	0.999974	-0.321988	
235	2	l	-0.014128	-0.014068	1.000000	-0.641533	
235	2	m	-0.403257	-0.402959	0.999978	-0.208961	
235	2	n	-0.434261	-0.433667	0.999974	-0.386604	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.401778	0.995723	1.285642	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209715	1.000000	1.908137	
248	1	j	-0.240205	-0.244108	0.999876	1.136981	
248	1	m	-0.384995	-0.400380	0.998027	1.172069	
248	1	n	-0.315971	-0.330202	0.998450	1.194390	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.410135	0.997319	1.398259	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.408083	0.998418	1.696249	
248	2	g	-0.385112	-0.407324	0.997400	1.462456	
248	2	h	-0.209574	-0.209576	1.000000	0.653858	
248	2	j	-0.253679	-0.255578	0.999934	0.764880	
248	2	l	-0.206419	-0.206449	1.000000	1.938001	
248	2	m	-0.426066	-0.440730	0.998495	1.294952	
248	2	n	-0.348954	-0.360402	0.998800	1.104669	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.082385	0.999928	1.929748	
252	1	d	0.146321	0.155483	0.997488	-0.554456	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.191518	0.998958	-0.364943	
252	1	j	0.209856	0.222480	0.994604	-0.527554	
252	1	m	0.200105	0.207138	0.998657	-0.587296	
252	1	n	0.253456	0.256671	0.999716	-0.590400	
252	2	a	0.006594	-0.000371	0.999896	2.045190	<****
252	2	b	-0.029094	-0.032597	0.999974	2.072719	<****
252	2	c	0.009133	0.002802	0.999913	2.039117	<****
252	2	d	0.188045	0.176063	0.998716	1.017375	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.076337	0.998618	1.514992	
252	2	g	0.196328	0.182969	0.998639	1.102349	
252	2	h	0.024025	0.024478	0.999987	-0.375015	
252	2	j	0.244992	0.242255	0.999147	0.289747	
252	2	l	0.000861	0.000896	1.000000	-0.374171	
252	2	m	0.248129	0.240101	0.999234	0.894038	
252	2	n	0.295229	0.293866	0.999794	0.297747	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.407137	1.000000	0.000000	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.182421	1.000000	0.000000	
222	1	j	-0.403934	-0.403934	1.000000	0.000000	
222	1	m	-0.430564	-0.430564	1.000000	0.000000	
222	1	n	-0.194080	-0.194080	1.000000	0.000000	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.410846	1.000000	0.000000	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.338840	1.000000	0.000000	
222	2	g	-0.410581	-0.410581	1.000000	0.000000	
222	2	h	-0.022308	-0.022308	1.000000	0.000000	
222	2	j	-0.430670	-0.430670	1.000000	0.000000	
222	2	l	-0.002454	-0.002454	1.000000	0.000000	
222	2	m	-0.442443	-0.442443	1.000000	0.000000	
222	2	n	-0.256865	-0.256865	1.000000	0.000000	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

PRO-VERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	b	-0.694873	-0.694873	1.000000	0.000000	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.373036	1.000000	0.000000	
223	1	j	-0.422509	-0.422509	1.000000	0.000000	
223	1	m	-0.695676	-0.695676	1.000000	0.000000	
223	1	n	-0.639829	-0.639829	1.000000	0.159475	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.732325	1.000000	0.000000	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.709220	1.000000	0.000000	
223	2	g	-0.731842	-0.731842	1.000000	0.000000	
223	2	h	-0.361158	-0.361158	1.000000	0.000000	
223	2	j	-0.389875	-0.389875	1.000000	0.000000	
223	2	l	-0.315803	-0.315803	1.000000	0.000000	
223	2	m	-0.726692	-0.726692	1.000000	0.000000	
223	2	n	-0.670259	-0.670259	1.000000	0.000000	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.166041	0.997118	1.413574	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.091571	0.996747	1.383275	
203	1	j	0.072857	0.060607	0.995188	0.545502	
203	1	m	-0.097276	-0.129761	0.995824	1.556263	
203	1	n	-0.012827	-0.055432	0.992866	1.556380	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.098771	0.997359	1.710038	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.064278	0.997835	1.857482	
203	2	g	-0.070989	-0.100535	0.997272	1.747608	
203	2	h	0.069005	0.056681	0.999148	1.303391	
203	2	j	0.034705	0.004534	0.996455	1.562676	
203	2	l	0.002329	0.001703	0.999998	1.358524	
203	2	m	-0.044468	-0.080896	0.996097	1.799743	
203	2	n	0.004985	-0.039334	0.993294	1.669367	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	-0.033285	0.991321	0.924409	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.207121	0.999440	0.913915	
207	1	j	0.168212	0.172674	0.999604	-0.718631	
207	1	m	0.095118	0.078800	0.996789	1.133186	
207	1	n	0.064855	0.041633	0.997764	1.554612	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	-0.000574	0.995388	1.048264	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.048246	0.999355	0.992676	
207	2	g	0.018805	-0.001071	0.996398	1.047422	
207	2	h	0.202720	0.195257	0.999464	1.036831	
207	2	j	0.193024	0.191117	0.999899	0.609649	
207	2	i	0.175987	0.167100	0.998359	0.703438	
207	2	m	0.121846	0.105408	0.998390	1.301823	
207	2	n	0.098820	0.079153	0.998625	1.681503	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.639311	0.998061	0.833893	
212	1	e	-0.281644	-0.326852	0.985549	1.354565	
212	1	h	0.089775	0.089848	1.000000	-1.045716	
212	1	j	-0.592539	-0.587989	0.999360	-0.762349	
212	1	m	-0.663786	-0.668711	0.998957	0.698407	
212	1	n	-0.701485	-0.698844	0.999350	-0.499063	
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.691929	0.997280	0.790613	
212	2	e	-0.288411	-0.322012	0.992112	1.362189	
212	2	f	-0.681729	-0.692550	0.997465	0.994981	
212	2	g	-0.692679	-0.701676	0.997236	0.810683	
212	2	h	0.055268	0.055271	1.000000	-0.027818	
212	2	j	-0.685346	-0.692573	0.999185	1.161042	
212	2	i	0.004343	0.004481	1.000000	-1.698082	
212	2	m	-0.710908	-0.718357	0.998150	0.892495	
212	2	n	-0.731150	-0.734941	0.998965	0.593250	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.435440	0.993732	0.168205	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.738434	0.999826	0.133021	
219	1	j	-0.671656	-0.666366	0.997785	-0.424835	
219	1	m	-0.541866	-0.539563	0.995675	-0.117736	
219	1	n	-0.562968	-0.558963	0.996748	-0.239634	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.396543	0.991506	0.035176	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.376363	0.991338	-0.139654	
219	2	g	-0.404803	-0.405428	0.991072	0.021454	
219	2	h	-0.769917	-0.770209	0.999729	0.078753	
219	2	j	-0.594641	-0.588325	0.996650	-0.331196	
219	2	l	-0.616435	-0.616045	0.999783	-0.094883	
219	2	m	-0.504693	-0.498291	0.994031	-0.270583	
219	2	n	-0.520418	-0.512288	0.994977	-0.377661	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	c	0.075409	0.052685	0.998553	1.582514	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.035029	1.000000	1.220531	
221	1	j	-0.072548	-0.100611	0.997579	1.512311	
221	1	m	0.012463	-0.012572	0.998193	1.558466	
221	1	n	-0.082838	-0.097667	0.999092	1.305427	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	0.033395	0.997325	0.357979	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.007429	0.996078	0.153394	
221	2	g	0.035177	0.028564	0.996834	0.310647	
221	2	h	-0.034210	-0.034089	1.000000	-1.221367	
221	2	j	-0.066140	-0.069308	0.999446	0.356998	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.015613	0.998669	0.153694	
221	2	n	-0.092854	-0.091717	0.999901	-0.303504	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.335791	0.958934	-1.186072	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.185837	0.999604	0.552102	
222	1	j	-0.403934	-0.371207	0.979887	-0.787117	
222	1	m	-0.430564	-0.371317	0.971151	-1.186905	
222	1	n	-0.194080	-0.198344	0.992970	0.164022	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.375819	0.981931	-0.889177	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.327114	0.989143	-0.377098	
222	2	g	-0.410581	-0.378757	0.982875	-0.831185	
222	2	h	-0.022308	-0.023060	0.999994	0.982573	
222	2	j	-0.430670	-0.397667	0.980397	-0.813171	
222	2	l	-0.002454	-0.002538	1.000000	0.648074	
222	2	m	-0.442443	-0.407506	0.979663	-0.848871	
222	2	n	-0.256865	-0.256679	0.990220	-0.006164	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.678547	0.960061	-0.298708	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.372791	0.999999	-0.870524	
223	1	j	-0.422509	-0.408334	0.999336	-1.584656	
223	1	m	-0.695676	-0.644752	0.976911	-1.126630	
223	1	n	-0.639829	-0.579433	0.974190	-1.159080	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.733488	0.978226	0.032851	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.717798	0.984534	0.262585	
223	2	g	-0.731842	-0.734684	0.979498	0.277675	
223	2	h	-0.361158	-0.359903	0.999975	-0.702043	
223	2	j	-0.389875	-0.384997	0.999839	-1.083664	
223	2	l	-0.315803	-0.315320	0.999997	-0.826624	
223	2	m	-0.726692	-0.704031	0.988673	-0.781386	
223	2	n	-0.670259	-0.639677	0.986943	-0.903943	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.364801	0.949885	0.654837	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.338044	0.990107	-0.157102	
227	1	j	0.059223	0.037574	0.968195	0.384391	
227	1	m	-0.227812	-0.268424	0.947537	0.578773	
227	1	n	-0.061523	-0.111317	0.960007	0.790844	
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.327784	0.966682	0.719152	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.322813	0.975858	0.862224	
227	2	g	-0.290104	-0.330734	0.967472	0.747528	
227	2	h	0.331148	0.336297	0.997306	-0.332534	
227	2	j	0.023736	0.002447	0.978523	0.459486	
227	2	l	0.037912	0.094589	0.985453	-1.489917	
227	2	m	-0.214012	-0.252393	0.964586	0.662769	
227	2	n	-0.066552	-0.111433	0.967368	0.789160	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	c	-0.251948	-0.245952	0.973601	-0.117476	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.076757	0.999979	0.008106	
230	1	j	0.058008	0.065870	0.997114	-0.451945	
230	1	m	-0.102239	-0.094741	0.985114	-0.190339	
230	1	n	-0.169797	-0.164029	0.983825	-0.141787	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000020	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.200564	0.977652	0.063798	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.156477	0.983113	0.165920	
230	2	g	-0.196159	-0.199755	0.976409	0.073618	
230	2	h	0.064555	0.068055	0.999842	-0.861271	
230	2	j	0.074391	0.081242	0.997614	-0.433525	
230	2	l	0.073664	0.074659	0.999982	-0.722802	
230	2	m	-0.075132	-0.077109	0.982990	0.046890	
230	2	n	-0.137167	-0.139718	0.980198	0.056424	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	d	-0.329375	-0.323510	0.998940	-0.570205	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.069413	0.995133	-1.062404	
235	1	j	-0.506407	-0.503594	0.997486	-0.194924	
235	1	m	-0.397157	-0.392297	0.998345	-0.389353	
235	1	n	-0.457890	-0.453587	0.997284	-0.277949	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.361747	0.998940	0.000722	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.417331	0.998649	0.430351	
235	2	g	-0.362648	-0.363351	0.998869	0.067272	
235	2	h	-0.126989	-0.120016	0.999415	-0.871020	
235	2	j	-0.453489	-0.455260	0.998212	0.140986	
235	2	l	-0.014128	-0.013328	0.999393	-0.906030	
235	2	m	-0.403257	-0.403586	0.998522	0.028018	
235	2	n	-0.434261	-0.434530	0.997962	0.019860	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999838	-0.990842	
248	1	d	-0.376824	-0.401778	0.995723	1.285642	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209715	1.000000	1.908137	
248	1	j	-0.240205	-0.244108	0.999876	1.136981	
248	1	m	-0.384995	-0.400380	0.998027	1.172069	
248	1	n	-0.315971	-0.330202	0.998450	1.194390	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.410135	0.997319	1.398259	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.408083	0.998418	1.696249	
248	2	g	-0.385112	-0.407324	0.997400	1.462456	
248	2	h	-0.209574	-0.209576	1.000000	0.653858	
248	2	j	-0.253679	-0.255578	0.999934	0.764880	
248	2	i	-0.206419	-0.206449	1.000000	1.938001	
248	2	m	-0.426066	-0.440730	0.998495	1.294952	
248	2	n	-0.348954	-0.360402	0.998800	1.104669	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

INDEFINITES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.082385	0.999928	1.929748	
252	1	d	0.146321	0.155483	0.997488	-0.554456	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.191518	0.998958	-0.364943	
252	1	j	0.209856	0.222480	0.994604	-0.527554	
252	1	m	0.200105	0.207138	0.998657	-0.587296	
252	1	n	0.253456	0.256671	0.999716	-0.590400	
252	2	a	0.006594	-0.000371	0.999896	2.045190	<****
252	2	b	-0.029094	-0.032597	0.999974	2.072719	<****
252	2	c	0.009133	0.002802	0.999913	2.039117	<****
252	2	d	0.188045	0.176063	0.998716	1.017375	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.076337	0.998618	1.514992	
252	2	g	0.196328	0.182969	0.998639	1.102349	
252	2	h	0.024025	0.024478	0.999987	-0.375015	
252	2	j	0.244992	0.242255	0.999147	0.289747	
252	2	i	0.000861	0.000896	1.000000	-0.374171	
252	2	m	0.248129	0.240101	0.999234	0.894038	
252	2	n	0.295229	0.293866	0.999794	0.297747	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.148431	0.998280	0.510848	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.109540	0.999843	1.856761	
203	1	j	0.072857	0.065661	0.998533	0.580401	
203	1	m	-0.097276	-0.107394	0.997989	0.698823	
203	1	n	-0.012827	-0.016290	0.998706	0.296710	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.082310	0.999104	1.237431	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.049983	0.999401	1.729432	
203	2	g	-0.070989	-0.084079	0.999124	1.365976	
203	2	h	0.069005	0.066236	0.999974	1.683621	
203	2	j	0.034705	0.022204	0.999062	1.258716	
203	2	i	0.002329	0.002250	1.000000	0.598162	
203	2	m	-0.044468	-0.058740	0.998959	1.364828	
203	2	n	0.004985	-0.002747	0.999308	0.905780	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	-0.006862	0.999871	0.222224	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.213823	1.000000	-0.147577	
207	1	j	0.168212	0.168294	0.999995	-0.119662	
207	1	m	0.093118	0.099079	0.999963	0.020531	
207	1	n	0.064855	0.065088	0.999974	-0.146065	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.022621	0.998963	-0.067669	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.038660	0.998391	-0.127885	
207	2	g	0.018805	0.019420	0.998886	-0.058332	
207	2	h	0.202720	0.202736	1.000000	-1.053751	
207	2	j	0.193024	0.193427	0.999992	-0.447260	
207	2	i	0.175987	0.176003	1.000000	-0.695023	
207	2	m	0.121846	0.122700	0.999803	-0.193783	
207	2	n	0.098820	0.100731	0.999934	-0.744773	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.617442	0.995464	-0.875176	
212	1	e	-0.281644	-0.311513	0.982702	0.821361	
212	1	h	0.089775	0.089747	1.000000	1.676563	
212	1	j	-0.592539	-0.580117	0.996792	-0.921043	
212	1	m	-0.663788	-0.650951	0.996615	-0.987932	
212	1	n	-0.701485	-0.686023	0.997651	-1.437058	
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.684623	0.997974	0.157717	
212	2	e	-0.288411	-0.299793	0.988173	0.379156	
212	2	f	-0.681729	-0.687163	0.998602	0.681362	
212	2	g	-0.692679	-0.694815	0.998031	0.231210	
212	2	h	0.055268	0.055208	1.000000	2.446004	(****
212	2	j	-0.685346	-0.685642	0.999039	0.045365	
212	2	l	0.004343	0.004343	1.000000	0.000000	
212	2	m	-0.710908	-0.711490	0.998727	0.080444	
212	2	n	-0.731150	-0.728661	0.999588	-0.613891	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.430073	0.999991	-1.145037	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.737903	1.000000	-1.709957	
219	1	j	-0.671656	-0.670533	0.999995	-1.692956	
219	1	m	-0.541866	-0.540698	0.999993	-1.370413	
219	1	n	-0.562968	-0.561691	0.999992	-1.422246	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.394188	0.999988	-1.154594	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.379951	0.999997	-1.141430	
219	2	g	-0.404803	-0.403586	0.999990	-1.151862	
219	2	h	-0.769917	-0.769824	1.000000	-1.823177	
219	2	j	-0.594641	-0.593935	0.999998	-1.560532	
219	2	l	-0.616435	-0.616371	1.000000	-1.472356	
219	2	m	-0.504693	-0.503757	0.999995	-1.360801	
219	2	n	-0.520418	-0.519481	0.999996	-1.397875	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	0.055584	0.994688	0.721152	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.034861	1.000000	-0.531102	
221	1	j	-0.072548	-0.084293	0.998980	0.975130	
221	1	m	0.012463	-0.004406	0.997030	0.819064	
221	1	n	-0.082838	-0.090558	0.999444	0.868784	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	0.022728	0.992589	0.562132	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.020693	0.993711	0.563734	
221	2	g	0.035177	0.016157	0.991892	0.559140	
221	2	h	-0.034210	-0.034158	1.000000	-0.529894	
221	2	j	-0.066140	-0.076286	0.998430	0.678972	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.027681	0.996252	0.613142	
221	2	n	-0.092854	-0.099614	0.999349	0.703887	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.402087	0.999147	-0.595397	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.182265	1.000000	-1.156142	
222	1	j	-0.403934	-0.402457	0.999895	-0.495627	
222	1	m	-0.430564	-0.427582	0.999605	-0.523476	
222	1	n	-0.194080	-0.195372	0.999992	1.506461	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.404633	0.999229	-0.769249	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.330964	0.999132	-0.891441	
222	2	g	-0.410581	-0.404013	0.999173	-0.784552	
222	2	h	-0.022308	-0.022308	1.000000	0.000000	
222	2	j	-0.430670	-0.426753	0.999656	-0.733350	
222	2	l	-0.002454	-0.002454	1.000000	0.000000	
222	2	m	-0.442443	-0.437722	0.999463	-0.712356	
222	2	n	-0.256865	-0.256762	0.999994	-0.132990	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.694671	0.999577	-0.036191	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.373034	1.000000	-0.040247	
223	1	j	-0.422509	-0.422606	0.999992	0.099842	
223	1	m	-0.695676	-0.694873	0.999624	-0.152344	
223	1	n	-0.639829	-0.639024	0.999351	-0.108592	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.731614	0.999758	-0.177229	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.708001	0.999524	-0.235080	
223	2	g	-0.731842	-0.730932	0.999726	-0.212666	
223	2	h	-0.361158	-0.361152	1.000000	-0.328827	
223	2	j	-0.389875	-0.389766	0.999997	-0.183723	
223	2	l	-0.315803	-0.315803	1.000000	0.000000	
223	2	.m	-0.726692	-0.725828	0.999753	-0.211110	
223	2	n	-0.670259	-0.668484	0.999639	-0.331034	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.376555	0.982406	1.385905	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.337958	0.990110	-0.154236	
227	1	j	0.059223	0.028478	0.993779	1.234047	
227	1	m	-0.227812	-0.282646	0.981532	1.310515	
227	1	n	-0.061523	-0.111850	0.982482	1.207144	
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.340908	0.984912	1.410279	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.331979	0.984462	1.310522	
227	2	g	-0.290104	-0.344101	0.984230	1.414211	
227	2	h	0.331148	0.336407	0.997310	-0.339819	
227	2	j	0.023736	-0.009689	0.992821	1.248016	
227	2	l	0.037912	0.094589	0.985453	-1.489917	
227	2	m	-0.214012	-0.265246	0.984349	1.325427	
227	2	n	-0.066552	-0.115033	0.984419	1.233237	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	d	-0.251948	-0.228853	0.953384	-0.339792	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.102748	0.995429	-1.188443	
230	1	j	0.058008	0.071562	0.996303	-0.688479	
230	1	m	-0.102239	-0.079787	0.975782	-0.446583	
230	1	n	-0.169797	-0.140125	0.973112	-0.564495	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.196111	0.961831	-0.022878	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.154653	0.970621	0.092608	
230	2	g	-0.196159	-0.196062	0.959678	-0.001521	
230	2	h	0.064555	0.083648	0.997631	-1.211830	
230	2	j	0.074391	0.075794	0.997745	-0.091302	
230	2	l	0.073664	0.080383	0.999684	-1.168519	
230	2	m	-0.075132	-0.075789	0.972691	0.012299	
230	2	n	-0.137167	-0.134730	0.969785	-0.043617	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	d	-0.329375	-0.328414	0.999977	-0.630002	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.094037	1.000000	-0.671878	
235	1	j	-0.506407	-0.504925	0.999937	-0.644094	
235	1	m	-0.397157	-0.395828	0.999960	-0.681098	
235	1	n	-0.457890	-0.456009	0.999933	-0.766970	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.360974	0.999981	-0.563587	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.412119	0.999994	-0.529427	
235	2	g	-0.362648	-0.361956	0.999984	-0.559762	
235	2	h	-0.126989	-0.126878	1.000000	-0.649706	
235	2	j	-0.453489	-0.452531	0.999976	-0.650605	
235	2	l	-0.014128	-0.014068	1.000000	-0.641533	
235	2	m	-0.403257	-0.402397	0.999981	-0.634816	
235	2	n	-0.434261	-0.433189	0.999975	-0.713140	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.432920	0.961978	0.986158	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209635	0.999999	-0.249116	
248	1	j	-0.240205	-0.243935	0.999592	0.601316	
248	1	m	-0.384995	-0.378511	0.985321	-0.163179	
248	1	n	-0.315971	-0.320346	0.991109	0.154769	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.458035	0.950784	1.079765	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.464474	0.932274	1.021533	
248	2	g	-0.385112	-0.457215	0.945090	1.062198	
248	2	h	-0.209574	-0.209490	0.999936	-0.137185	
248	2	j	-0.253679	-0.246109	0.998955	-0.762547	
248	2	l	-0.206419	-0.208055	0.998351	0.130231	
248	2	m	-0.426066	-0.402240	0.964065	-0.436521	
248	2	n	-0.348954	-0.337148	0.974595	-0.249442	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

RESIDUAL ADJECTIVES

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.142479	0.999663	0.633625	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.192027	0.998954	-0.412223	
252	1	j	0.209856	0.220785	0.995109	-0.479707	
252	1	m	0.200105	0.198632	0.999614	0.229734	
252	1	n	0.253456	0.252892	0.999964	0.292842	
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	d	0.188045	0.182595	0.999405	0.681319	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.089059	0.999268	0.668736	
252	2	g	0.196328	0.190393	0.999344	0.707253	
252	2	h	0.024025	0.024478	0.999987	-0.375015	
252	2	j	0.244992	0.246097	0.999414	-0.141210	
252	2	l	0.000861	0.000896	1.000000	-0.374171	
252	2	m	0.248129	0.244924	0.999630	0.514756	
252	2	n	0.295229	0.294767	0.999967	0.250804	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.142082	0.999987	0.396757	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.117064	1.000000	0.000000	
203	1	j	0.072857	0.072786	1.000000	0.411022	
203	1	m	-0.097276	-0.097666	0.999994	0.513715	
203	1	n	-0.012827	-0.013044	0.999999	0.849521	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.070961	0.999984	0.501539	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.036557	0.999997	0.535096	
203	2	g	-0.070989	-0.071555	0.999988	0.496935	
203	2	h	0.069005	0.069005	1.000000	0.000000	
203	2	j	0.034705	0.034300	0.999995	0.544439	
203	2	l	0.002329	0.002329	1.000000	0.000000	
203	2	m	-0.044468	-0.045014	0.999990	0.526024	
203	2	n	0.004985	0.004512	0.999996	0.710858	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine, #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	0.032065	0.992816	-1.423458	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.213874	1.000000	-0.850277	
207	1	j	0.168212	0.174635	0.999839	-1.615627	
207	1	m	0.099118	0.121884	0.997082	-1.338973	
207	1	n	0.064855	0.084285	0.997783	-1.307989	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	0.048181	0.997133	-1.551208	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.025129	0.998944	-1.475142	
207	2	g	0.018805	0.042610	0.997630	-1.547063	
207	2	h	0.202720	0.202787	1.000000	-1.210457	
207	2	j	0.193024	0.195876	0.999971	-1.702039	
207	2	l	0.175987	0.176039	1.000000	-1.043860	
207	2	m	0.121846	0.135433	0.999145	-1.477537	
207	2	n	0.098820	0.111597	0.999273	-1.504133	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.640823	0.997044	0.799894	
212	1	e	-0.281644	-0.331317	0.984298	1.427274	
212	1	h	0.089775	0.088470	0.999987	1.250839	
212	1	j	-0.592539	-0.604145	0.997900	1.067061	
212	1	m	-0.663788	-0.677385	0.998447	1.505831	
212	1	n	-0.701485	-0.715708	0.999080	1.994759	(****
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.679519	0.995743	-0.261418	
212	2	e	-0.288411	-0.327933	0.991660	1.554373	
212	2	f	-0.681729	-0.676123	0.995372	-0.387514	
212	2	g	-0.692679	-0.688354	0.995545	-0.309736	
212	2	h	0.055268	0.052574	0.999918	1.030669	
212	2	j	-0.685346	-0.684692	0.998036	-0.070164	
212	2	l	0.004343	0.004002	0.999999	1.006818	
212	2	m	-0.710908	-0.707554	0.996714	-0.286978	
212	2	n	-0.731150	-0.726278	0.997792	-0.520132	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331750	0.999978	0.212136	
219	1	d	-0.431196	-0.457892	0.993178	1.002152	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.728622	0.999181	-1.252895	
219	1	j	-0.671656	-0.679750	0.999325	1.128190	
219	1	m	-0.541866	-0.557848	0.997718	1.094260	
219	1	n	-0.562968	-0.578395	0.998266	1.217319	
219	2	a	-0.253123	-0.254585	0.999933	0.519852	
219	2	b	-0.223322	-0.224152	0.999984	0.606919	
219	2	c	-0.264359	-0.265601	0.999944	0.486586	
219	2	d	-0.395491	-0.423484	0.992759	1.004290	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.409756	0.991021	0.935396	
219	2	g	-0.404803	-0.432203	0.992215	0.953313	
219	2	h	-0.769917	-0.767515	0.999397	-0.428214	
219	2	j	-0.594641	-0.600761	0.999318	0.808325	
219	2	l	-0.616435	-0.615349	0.999353	-0.153117	
219	2	m	-0.504693	-0.518620	0.997799	0.955362	
219	2	n	-0.520418	-0.532008	0.998432	0.950606	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	0.068418	0.999671	1.021262	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.034912	1.000000	0.000000	
221	1	j	-0.072548	-0.078012	0.999855	1.204255	
221	1	m	0.012463	0.006072	0.999766	1.105415	
221	1	n	-0.082838	-0.085460	0.999962	1.131846	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	0.039240	0.999922	0.524472	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.004296	0.999986	0.347925	
221	2	g	0.035177	0.033652	0.999936	0.505707	
221	2	h	-0.034210	-0.034210	1.000000	0.000000	
221	2	j	-0.066140	-0.067458	0.999979	0.754856	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.014859	0.999966	0.617133	
221	2	n	-0.092854	-0.093602	0.999993	0.738241	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	c	-0.407137	-0.407137	1.000000	0.000000	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.182421	1.000000	0.000000	
222	1	j	-0.403934	-0.403934	1.000000	0.000000	
222	1	m	-0.430564	-0.430564	1.000000	0.000000	
222	1	n	-0.194080	-0.194080	1.000000	0.000000	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	c	-0.410846	-0.410846	1.000000	0.000000	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.338840	1.000000	0.000000	
222	2	g	-0.410581	-0.410581	1.000000	0.000000	
222	2	h	-0.022308	-0.022308	1.000000	0.000000	
222	2	j	-0.430670	-0.430670	1.000000	0.000000	
222	2	i	-0.002454	-0.002454	1.000000	0.000000	
222	2	m	-0.442443	-0.442443	1.000000	0.000000	
222	2	n	-0.256865	-0.256865	1.000000	0.000000	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.689378	0.999583	-0.940315	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.373061	1.000000	0.175582	
223	1	j	-0.422509	-0.427477	0.999156	0.497421	
223	1	m	-0.695676	-0.699367	0.998016	0.304605	
223	1	n	-0.639829	-0.646247	0.997095	0.408513	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.729744	0.999907	-0.979208	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.707628	0.999940	-0.745860	
223	2	g	-0.731842	-0.729464	0.999921	-0.979905	
223	2	h	-0.361158	-0.361454	0.999987	0.231277	
223	2	j	-0.389875	-0.391907	0.999764	0.379268	
223	2	l	-0.315803	-0.316129	0.999994	0.357899	
223	2	m	-0.726692	-0.729550	0.999522	0.497603	
223	2	n	-0.670259	-0.674251	0.999297	0.530942	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	p > .05
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.398033	0.972933	1.550742	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.335310	0.990081	-0.064722	
227	1	j	0.059223	-0.073285	0.967392	2.334452	(****
227	1	m	-0.227812	-0.334139	0.967161	1.901058	
227	1	n	-0.061523	-0.214835	0.948281	2.156459	(****
227	2	a	-0.122705	-0.122705	1.000000	0.000000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.350143	0.978771	1.400076	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.333198	0.980243	1.193629	
227	2	g	-0.290104	-0.351318	0.978304	1.370179	
227	2	h	0.331148	0.332918	0.997930	-0.131987	
227	2	j	0.023736	-0.085362	0.975891	2.231344	(****
227	2	l	0.037912	0.089832	0.987456	-1.469408	
227	2	m	-0.214012	-0.297705	0.975253	1.720195	
227	2	n	-0.066552	-0.199129	0.956814	2.037332	(****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

Page

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	d	-0.251948	-0.269100	0.996351	0.902719	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.076110	0.999996	1.031122	
230	1	j	0.058008	0.044673	0.998714	1.147541	
230	1	m	-0.102239	-0.148149	0.978426	0.970751	
230	1	n	-0.169797	-0.198683	0.990710	0.937975	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.206214	0.999202	0.964329	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.153486	0.999852	1.003149	
230	2	g	-0.196159	-0.204202	0.999332	0.975792	
230	2	h	0.064555	0.061481	0.999915	1.028799	
230	2	j	0.074391	0.062550	0.998378	0.908082	
230	2	l	0.073664	0.072332	0.999965	0.697923	
230	2	m	-0.075132	-0.092989	0.995770	0.849063	
230	2	n	-0.137167	-0.148515	0.998314	0.859593	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	c	-0.329375	-0.360922	0.997923	2.113196	(****
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.099305	0.998099	0.363568	
235	1	j	-0.506407	-0.546140	0.994535	1.758145	
235	1	m	-0.397157	-0.434698	0.996979	2.105674	(****
235	1	n	-0.457890	-0.494441	0.996343	1.915425	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.225216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.371986	0.999068	1.068200	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.409704	0.999308	-0.351968	
235	2	g	-0.362648	-0.369926	0.999186	0.816485	
235	2	h	-0.126989	-0.127378	0.999892	0.113158	
235	2	j	-0.453489	-0.472193	0.998358	1.498533	
235	2	l	-0.014128	-0.014163	0.999999	0.094714	
235	2	m	-0.403257	-0.414596	0.998859	1.085507	
235	2	n	-0.434261	-0.444998	0.998627	0.953844	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	p > .05
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376324	-0.405021	0.997801	1.972311	(****
246	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209715	1.000000	2.047406	(****
248	1	j	-0.240205	-0.245887	0.999913	1.947197	
248	1	m	-0.384995	-0.405509	0.998981	2.096243	(****
248	1	n	-0.315971	-0.335770	0.999108	2.135462	(****
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.150639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.414537	0.998551	2.206564	(****
246	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.410657	0.998757	2.113281	(****
248	2	g	-0.385112	-0.411221	0.998479	2.174912	(****
248	2	h	-0.209574	-0.209579	1.000000	1.967485	(****
248	2	j	-0.253679	-0.257323	0.999970	2.113583	(****
248	2	l	-0.206419	-0.206449	1.000000	1.938001	
248	2	m	-0.426066	-0.445533	0.999258	2.312804	(****
248	2	n	-0.348954	-0.366078	0.999461	2.361162	(****

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

ADVERBS

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.152037	0.999355	-0.682409	
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.191735	0.998957	-0.385191	
252	1	j	0.209856	0.222346	0.995179	-0.552108	
252	1	m	0.200105	0.204403	0.999706	-0.766240	
252	1	n	0.253456	0.253736	0.999994	-0.342848	
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	c	0.188045	0.188014	0.999973	0.017908	
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.092604	0.999813	0.542856	
252	2	g	0.196328	0.195842	0.999955	0.222607	
252	2	h	0.024025	0.024478	0.999987	-0.375015	
252	2	j	0.244992	0.248101	0.999548	-0.452515	
252	2	l	0.000861	0.000896	1.000000	-0.374171	
252	2	m	0.248129	0.248331	0.999918	-0.068750	
252	2	n	0.295229	0.294360	0.999923	0.311041	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
203	1	a	-0.105343	-0.106148	0.999980	0.553864	
203	1	d	-0.141626	-0.161493	0.991356	0.665910	
203	1	e	-0.001022	-0.001022	1.000000	0.000000	
203	1	h	0.117064	0.052089	0.985430	1.665079	
203	1	j	0.072857	-0.013138	0.989212	2.558135	<****
203	1	m	-0.097276	-0.148119	0.990305	1.600324	
203	1	n	-0.012827	-0.055864	0.994042	1.720215	
203	2	a	-0.042574	-0.043710	0.999964	0.582901	
203	2	b	0.005796	0.005370	0.999995	0.574551	
203	2	c	-0.037613	-0.038579	0.999974	0.585132	
203	2	d	-0.070321	-0.127871	0.981672	1.316604	
203	2	e	-0.001022	-0.001022	1.000000	0.000000	
203	2	f	-0.036263	-0.093124	0.981955	1.308198	
203	2	g	-0.070989	-0.130450	0.980511	1.319463	
203	2	h	0.069005	0.048481	0.995322	0.926427	
203	2	j	0.034705	-0.052601	0.978173	1.826351	
203	2	l	0.002329	0.001232	0.999989	1.031834	
203	2	m	-0.044468	-0.129594	0.976002	1.701714	
203	2	n	0.004985	-0.059381	0.984518	1.597165	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
207	1	a	0.089728	0.088733	0.999863	0.269908	
207	1	d	-0.006064	-0.077743	0.962852	1.179090	
207	1	e	-0.000956	-0.000956	1.000000	0.000000	
207	1	h	0.213822	0.220392	0.999440	-0.896748	
207	1	j	0.168212	0.187751	0.998075	-1.424611	
207	1	m	0.099118	0.083089	0.987516	0.455556	
207	1	n	0.064855	0.041813	0.977726	0.489005	
207	2	a	0.069602	0.068528	0.999717	0.202447	
207	2	b	0.054739	0.054262	0.999936	0.188977	
207	2	c	0.070263	0.069245	0.999769	0.212202	
207	2	d	0.021932	-0.017484	0.971877	0.743697	
207	2	e	-0.000956	-0.000956	1.000000	0.000000	
207	2	f	-0.040281	-0.047937	0.971695	0.144053	
207	2	g	0.018805	-0.014957	0.972672	0.646113	
207	2	h	0.202720	0.214014	0.998776	-1.040205	
207	2	j	0.193024	0.202776	0.999022	-1.002807	
207	2	l	0.175987	0.203602	0.979143	-0.615540	
207	2	m	0.121846	0.117138	0.989392	0.145574	
207	2	n	0.098820	0.082496	0.983386	0.402153	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
212	1	a	-0.532393	-0.534533	0.999965	1.423618	
212	1	d	-0.630979	-0.638314	0.992509	0.379011	
212	1	e	-0.281644	-0.380416	0.964098	1.876376	
212	1	h	0.089775	0.084509	0.999889	1.732290	
212	1	j	-0.592539	-0.561817	0.993563	-1.543675	
212	1	m	-0.663788	-0.650203	0.994932	-0.860060	
212	1	n	-0.701485	-0.660807	0.993845	-2.115928	<****
212	2	a	-0.553674	-0.556193	0.999956	1.510422	
212	2	b	-0.536118	-0.537247	0.999990	1.407011	
212	2	c	-0.560078	-0.562278	0.999967	1.525915	
212	2	d	-0.683127	-0.663560	0.992462	-1.025839	
212	2	e	-0.288411	-0.394990	0.969909	2.195630	<****
212	2	f	-0.681729	-0.664523	0.990567	-0.815818	
212	2	g	-0.692679	-0.671592	0.992099	-1.086799	
212	2	h	0.055268	0.051919	0.999929	1.379969	
212	2	j	-0.685346	-0.655242	0.994465	-1.726480	
212	2	l	0.004343	0.003938	0.999999	1.336510	
212	2	m	-0.710908	-0.682914	0.993353	-1.538147	
212	2	n	-0.731150	-0.693535	0.993167	-1.959042	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
219	1	a	-0.331415	-0.331415	1.000000	0.000000	
219	1	d	-0.431196	-0.458912	0.994957	1.200153	
219	1	e	-0.001101	-0.001101	1.000000	0.000000	
219	1	h	-0.738016	-0.739108	0.999833	0.352359	
219	1	j	-0.671656	-0.674625	0.998730	0.317383	
219	1	m	-0.541866	-0.555294	0.996965	0.809469	
219	1	n	-0.562968	-0.576185	0.997133	0.831686	
219	2	a	-0.253123	-0.253123	1.000000	0.000000	
219	2	b	-0.223322	-0.223322	1.000000	0.000000	
219	2	c	-0.264359	-0.264359	1.000000	0.000000	
219	2	d	-0.395491	-0.411132	0.992687	0.562857	
219	2	e	-0.001101	-0.001101	1.000000	0.000000	
219	2	f	-0.380616	-0.384239	0.991255	0.118603	
219	2	g	-0.404803	-0.419032	0.992614	0.512025	
219	2	h	-0.769917	-0.770346	0.999754	0.121066	
219	2	j	-0.594641	-0.595432	0.997550	0.056217	
219	2	l	-0.616435	-0.616531	0.999783	0.023530	
219	2	m	-0.504693	-0.510413	0.995036	0.266190	
219	2	n	-0.520418	-0.526431	0.995527	0.297916	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
221	1	a	0.044923	0.044756	0.999961	0.071199	
221	1	d	0.075409	0.085339	0.972765	-0.159731	
221	1	e	-0.000885	-0.000885	1.000000	0.000000	
221	1	h	-0.034912	-0.031620	0.999956	-1.311737	
221	1	j	-0.072548	-0.048756	0.989701	-0.621482	
221	1	m	0.012463	0.033271	0.980126	-0.390667	
221	1	n	-0.082838	-0.054731	0.991137	-0.791764	
221	2	a	0.035569	0.035454	0.999933	0.037527	
221	2	b	0.024068	0.023727	0.999973	0.173684	
221	2	c	0.035314	0.035149	0.999942	0.057249	
221	2	d	0.040988	0.021364	0.965358	0.279127	
221	2	e	-0.000885	-0.000885	1.000000	0.000000	
221	2	f	-0.003799	-0.038791	0.969275	0.528531	
221	2	g	0.035177	0.010157	0.962259	0.340917	
221	2	h	-0.034210	-0.032175	0.999985	-1.368177	
221	2	j	-0.066140	-0.057768	0.989497	-0.216570	
221	2	l	-0.001531	-0.001531	1.000000	0.000000	
221	2	m	-0.013498	-0.021055	0.980327	0.142565	
221	2	n	-0.092854	-0.082137	0.994236	-0.374885	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
222	1	a	-0.065303	-0.065303	1.000000	0.000000	
222	1	d	-0.407137	-0.431149	0.950314	0.375593	
222	1	e	-0.000884	-0.000884	1.000000	0.000000	
222	1	h	-0.182421	-0.203050	0.999068	2.149988	(****
222	1	j	-0.403934	-0.428106	0.954801	0.395615	
222	1	m	-0.430564	-0.459139	0.950352	0.452999	
222	1	n	-0.194080	-0.197124	0.986310	0.083897	
222	2	a	-0.076259	-0.076259	1.000000	0.000000	
222	2	b	-0.063133	-0.063133	1.000000	0.000000	
222	2	c	-0.073564	-0.073564	1.000000	0.000000	
222	2	d	-0.410846	-0.467872	0.951402	0.903768	
222	2	e	-0.000884	-0.000884	1.000000	0.000000	
222	2	f	-0.338840	-0.433130	0.951294	1.438799	
222	2	g	-0.410581	-0.473168	0.952301	0.999968	
222	2	h	-0.022308	-0.022024	0.999931	-0.108190	
222	2	j	-0.430670	-0.482421	0.961483	0.928051	
222	2	l	-0.002454	-0.002454	1.000000	0.000070	
222	2	m	-0.442443	-0.503002	0.951921	0.980138	
222	2	n	-0.256865	-0.296653	0.984493	1.045795	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
223	1	a	-0.732594	-0.732594	1.000000	0.000000	
223	1	d	-0.694873	-0.597237	0.909762	-1.089483	
223	1	e	-0.000648	-0.000648	1.000000	0.000000	
223	1	h	-0.373036	-0.413896	0.995578	1.684466	
223	1	j	-0.422509	-0.450678	0.993062	0.975742	
223	1	m	-0.695676	-0.612015	0.944864	-1.181094	
223	1	n	-0.639829	-0.543172	0.934936	-1.192056	
223	2	a	-0.754136	-0.754136	1.000000	0.000000	
223	2	b	-0.718551	-0.718551	1.000000	0.000000	
223	2	c	-0.752835	-0.752835	1.000000	0.000000	
223	2	d	-0.732325	-0.654944	0.930620	-1.038573	
223	2	e	-0.000648	-0.000648	1.000000	0.000000	
223	2	f	-0.709220	-0.658562	0.945831	-0.775813	
223	2	g	-0.731842	-0.655799	0.932187	-1.032759	
223	2	h	-0.351158	-0.383670	0.998772	1.747008	
223	2	j	-0.389875	-0.401449	0.998356	0.813411	
223	2	l	-0.315803	-0.324859	0.999795	1.707377	
223	2	m	-0.726692	-0.667592	0.954746	-0.985135	
223	2	n	-0.670259	-0.586862	0.943080	-1.137525	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
227	1	a	-0.128573	-0.128573	1.000000	0.000000	
227	1	d	-0.321199	-0.377591	0.925556	0.697176	
227	1	e	-0.000658	-0.000658	1.000000	0.000000	
227	1	h	0.333390	0.338081	0.990107	-0.158368	
227	1	j	0.059223	-0.001564	0.953721	0.895110	
227	1	m	-0.227812	-0.292655	0.911602	0.714758	
227	1	n	-0.061523	-0.123365	0.929369	0.739915	
227	2	a	-0.122705	-0.122705	1.000000	0.200000	
227	2	b	-0.133699	-0.133699	1.000000	0.000000	
227	2	c	-0.126353	-0.126353	1.000000	0.000000	
227	2	d	-0.288205	-0.354688	0.961845	1.128635	
227	2	e	-0.000658	-0.000658	1.000000	0.000000	
227	2	f	-0.282278	-0.336992	0.974451	1.130055	
227	2	g	-0.290104	-0.356649	0.962468	1.139455	
227	2	h	0.331148	0.336556	0.997301	-0.348911	
227	2	j	0.023736	-0.025261	0.972956	0.943011	
227	2	l	0.037912	0.094668	0.985453	-1.492026	
227	2	m	-0.214012	-0.277742	0.956547	0.995203	
227	2	n	-0.066552	-0.120247	0.961141	0.865730	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
230	1	a	-0.373570	-0.373570	1.000000	0.000000	
230	1	d	-0.251948	-0.232470	0.938043	-0.248792	
230	1	e	-0.000600	-0.000600	1.000000	0.000000	
230	1	h	0.076769	0.075902	0.999936	0.334578	
230	1	j	0.058008	0.045139	0.992559	0.460424	
230	1	m	-0.102239	-0.132519	0.924660	0.342555	
230	1	n	-0.169797	-0.182024	0.939215	0.155354	
230	2	a	-0.305620	-0.305620	1.000000	0.000000	
230	2	b	-0.234657	-0.234657	1.000000	0.000000	
230	2	c	-0.301274	-0.301274	1.000000	0.000000	
230	2	d	-0.197532	-0.187890	0.943602	-0.127603	
230	2	e	-0.000600	-0.000600	1.000000	0.000000	
230	2	f	-0.149564	-0.141701	0.947459	-0.106907	
230	2	g	-0.196159	-0.185741	0.942992	-0.137083	
230	2	h	0.064555	0.060895	0.999806	0.810542	
230	2	j	0.074391	0.060171	0.994686	0.602574	
230	2	l	0.073664	0.071688	0.999834	0.473614	
230	2	m	-0.075132	-0.094402	0.953716	0.277135	
230	2	n	-0.137167	-0.149996	0.957739	0.194410	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
235	1	a	-0.031123	-0.031123	1.000000	0.000000	
235	1	c	-0.329375	-0.285752	0.994152	-1.755836	
235	1	e	-0.000931	-0.000931	1.000000	0.000000	
235	1	h	-0.094045	-0.057994	0.987647	-0.975960	
235	1	j	-0.506407	-0.483021	0.996102	-1.251740	
235	1	m	-0.397157	-0.362934	0.993628	-1.360246	
235	1	n	-0.457890	-0.430937	0.993811	-1.124012	
235	2	a	-0.160684	-0.160684	1.000000	0.000000	
235	2	b	-0.226216	-0.226216	1.000000	0.000000	
235	2	c	-0.155708	-0.155708	1.000000	0.000000	
235	2	d	-0.361740	-0.309087	0.991218	-1.740057	
235	2	e	-0.000931	-0.000931	1.000000	0.000000	
235	2	f	-0.412523	-0.358901	0.985995	-1.437685	
235	2	g	-0.362648	-0.308408	0.990090	-1.690505	
235	2	h	-0.126989	-0.117422	0.998917	-0.877817	
235	2	j	-0.453489	-0.413285	0.992495	-1.491843	
235	2	l	-0.014128	-0.013093	0.999987	-0.872076	
235	2	m	-0.423257	-0.351433	0.989886	-1.619075	
235	2	n	-0.434261	-0.386990	0.990077	-1.512486	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
248	1	a	-0.174181	-0.171062	0.999898	-0.990842	
248	1	d	-0.376824	-0.378582	0.955880	0.028650	
248	1	e	-0.001000	-0.001000	1.000000	0.000000	
248	1	h	-0.209710	-0.209676	1.000000	-0.427821	
248	1	j	-0.240205	-0.256509	0.997028	0.972292	
248	1	m	-0.384995	-0.414237	0.966508	0.550285	
248	1	n	-0.315971	-0.353847	0.975395	0.807042	
248	2	a	-0.162555	-0.158754	0.999855	-1.010935	
248	2	b	-0.162328	-0.160639	0.999969	-0.978255	
248	2	c	-0.165652	-0.162231	0.999882	-1.006056	
248	2	d	-0.388644	-0.365281	0.939041	-0.323601	
248	2	e	-0.001000	-0.001000	1.000000	0.000000	
248	2	f	-0.387811	-0.363248	0.932480	-0.323157	
248	2	g	-0.385112	-0.358866	0.934999	-0.351281	
248	2	h	-0.209574	-0.209563	1.000000	-0.157523	
248	2	j	-0.253679	-0.262218	0.997347	0.541879	
248	2	l	-0.206419	-0.206416	1.000000	-0.022628	
248	2	m	-0.426066	-0.422344	0.952375	-0.059719	
248	2	n	-0.348954	-0.359561	0.968009	0.200703	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

A Statistical Comparison of the Relationship Between Unresolved Anaphors and User's Relevance Judgments with Resolved Anaphors and User's Relevance Judgments: for Anaphoric Class

DEFINITE ARTICLE

			Correlation Coefficients			Significance Level	
Q	S	TW	r_{ju}	r_{jr}	r_{ur}	Z	$p > .05$
252	1	a	0.087821	0.087821	1.000000	0.000000	
252	1	d	0.146321	0.283660	0.971175	-2.443995	<****
252	1	e	-0.000931	-0.000931	1.000000	0.000000	
252	1	h	0.187661	0.229938	0.989385	-1.252583	
252	1	j	0.209856	0.328685	0.980117	-2.534868	<****
252	1	m	0.200105	0.325146	0.976233	-2.448122	<****
252	1	n	0.253456	0.326274	0.987219	-1.968003	<****
252	2	a	0.006594	0.006594	1.000000	0.000000	
252	2	b	-0.029094	-0.029094	1.000000	0.000000	
252	2	c	0.009133	0.009133	1.000000	0.000000	
252	2	d	0.188045	0.341037	0.956129	-2.226513	<****
252	2	e	-0.000931	-0.000931	1.000000	0.000000	
252	2	f	0.095067	0.282277	0.927175	-2.118646	<****
252	2	g	0.196328	0.355381	0.949463	-2.164196	<****
252	2	h	0.024025	0.029704	0.999870	-1.493036	
252	2	j	0.244992	0.352520	0.977306	-2.174153	<****
252	2	l	0.000861	0.001121	1.000000	-1.124364	
252	2	m	0.248129	0.375791	0.966095	-2.122799	<****
252	2	n	0.295229	0.358758	0.987346	-1.746842	

NOTES:

Q: Queries 100-199 were searched on INSPEC: 200-299 on PsychINFO

S: Similarity Measure: #1 = Cosine. #2 = Dice

TW: Term Weighting Schemes: See Result Page R-1

Correlation Coefficients: r_{ju} is between the user's relevance judgment and the system's predicted relevance based on unresolved anaphors. r_{jr} is between the user's relevance judgment and the system's predicted relevance based on resolved anaphors.

Because the user's judgments were scaled from low to high (1 = most relevant, 4 = most non-relevant) a strong negative correlation shows agreement between user's and system's relevance judgments.

Significance Level: A positive Z indicates that the second correlation is higher than the first correlation ($r_{jr} > r_{ju}$). If this Z is statistically significant as indicated by the asterisks, then resolving anaphors improves the system's predications of relevance.

APPENDIX F

Summaries of
Statistical Results,
INSPEC and PsycABS

Summary of Statistical Results
By INSPEC Query

INSPEC Query	Similarity Measure and Term Weighting																			Total + /-
	1A	1D	1E	1H	1J	1M	1N	2A	2B	2C	2D	2E	2F	2G	2H	2J	2L	2M	2N	
101																			-J	0/-1
103		C -J				C -J	C -J						C	C				C	C	7/-3
104					-J											-J		-J	-J	0/-4
107																J R				2/-0
109					-B -F	-G	-G								-B	-E -G		-G	-G	0/-9
135							-F													0/-1
142																				0/0
158															D,E, F,G, H,I					6/0
170		-R -B			-B	-R -B	-B				-B		-B	-B	J	-B		-B	-B J	2/-12
180				R							F		D F	F				E,F, H,I		9/0
182																				0/0
184																				0/0
Total + /-	0/0	1/-3	0/0	1/0	0/-4	1/-4	1/-4	0/0	0/0	0/0	1/-1	0/0	3/-1	2/-1	7/-1	2/-4	0 0	5 -3	2 -4	26/-30

(Cell entries indicate the class of anaphor producing a statistically significant finding

Negative sign indicates that resolution decreases retrieval performance.

Summary of Statistical Results
- By Psychological Abstracts Query

PsycINFO Query	Similarity Measure and Term Weighting																			Total + /-
	1A	1B	1E	1H	1J	1M	1N	2A	2B	2C	2D	2E	2F	2G	2H	2J	2L	2M	2N	
203	I,R	C		I,R	I,J	I	I	I,R	I,R	I,R	I,R		I,R	I,R	I	I,R	I	I,R	I,R	29 /0
207				A											A					2 /0
212			A,B, C				H,-J -R					A,B, C,J, R			G				-R	10 /-3
219		D			D						C		C	C				C	C,D	8 /0
221	-I	A,R			A	A	A	-I	-I	-I										5 /-4
222	-R			-J				-R	-R	-R										0 /-5
223																				0 /0
227					H	H										H			H	4 /0
230		-B -C				-B	-B				-B -C		-C	-B -C				-B	-B -C	0 /-12
235		H		B		H														3 /0
248	-E	H		H	B	H	H	-E	-E	-E	H		H	H	H	H		H	H	12 /-4
252		-C,J -R			-C,J -R	-C,J	J	E,F	E,F	E,F	J,-R		J,-R	J,-R	-C	J,-R		J		15 /-10
Total + /-	2/-3	7/-4	3/0	5/-1	7/-2	5/-2	6/-3	4/-3	4/-3	4/-3	5/-3	5/0	5/-2	5/-3	4/-1	5/-1	1/0	5/-1	6/-3	88 /-38

(Cell entries indicate the class of anaphor producing a statistically significant finding
Negative sign indicates that resolution decreases retrieval performance.)

Summary of Statistical Results
By INSPEC Query

INSPEC Query	Similarity Measure and Term Weighting																			Total + /-
	1A	1D	1E	1H	1J	1M	1N	2A	2B	2C	2D	2E	2F	2G	2H	2J	2L	2M	2N	
101																			-J	0/-1
103		C -J				C -J	C -J						C	C				C	C	7/-3
104					-J											-J		-J	-J	0/-4
107																J R				2/-0
109					-B -F	-G	-G								-B	-E -G		-G	-G	0/-9
135							-F													0/-1
142																				0/0
158															D,E, F,G, H,I					6/0
170		-R -B			-B	-R -B	-B				-B		-B	-B	J	-B		-B	-B J	2/-12
180				R							F		D F	F				E,F, H,I		9/0
182																				0/0
184																				0/0
Total + /-	0/0	1/-3	0/0	1/0	0/-4	1/-4	1/-4	0/0	0/0	0/0	1/-1	0/0	3/-1	2/-1	7/-1	2/-4	0 0	5 -3	2 -4	26/-30

(Cell entries indicate the class of anaphor producing a statistically significant finding

Negative sign indicates that resolution decreases retrieval performance.

Summary of Statistical Results
By Psychological Abstracts Anaphoric Class

Anaphoric Class	Similarity Measure and Term Weighting																			Total
	1A	1D	1E	1H	1J	1M	1N	2A	2B	2C	2D	2E	2F	2G	2H	2J	2L	2M	2N	+ /-
A		1	1	1	1	1	1								1					7 /-0
B		-1	1	1	1	-1	-1				-1			-1				-1	-1	3 /-7
C		1_{-2}	1		-1	-1					1_{-1}		1_{-1}	1_{-1}	-1			1	1_{-1}	7 /-9
D		1			1														1	3 /-0
E	-1							1_{-1}	1_{-1}	1_{-1}										3 /-4
F								1	1	1										3 /-0
G															1					1 /-0
H		2		1	1	2	3				1		1	1	1	2		1	2	18 /-0
I	1_{-1}			1	1	1	1	1_{-1}	1_{-1}	1_{-1}	1		1	1	1	1	1	1	1	16 /-4
J		1		-1	2	1	1_{-1}				1		1	1		1		1		10 /-2
R	1_{-1}	1_{-1}		1	-1		-1	1_{-1}	1_{-1}	1_{-1}	1_{-1}		1_{-1}	1_{-1}		1_{-1}		1	1_{-1}	12 /-12
Total + /-	2/-3	7/-4	3/0	5/-1	7/-2	5/-2	6/-3	4/-3	4/-3	4/-3	5/-3	0/0	5/-2	5/-3	4/-1	5/-1	1/0	5/-1	6/-3	83 /-38

(Cell entries contain the number of PsycINFO queries with statistically significant findings.

Negative sign indicates that resolution decreases retrieval performance.)